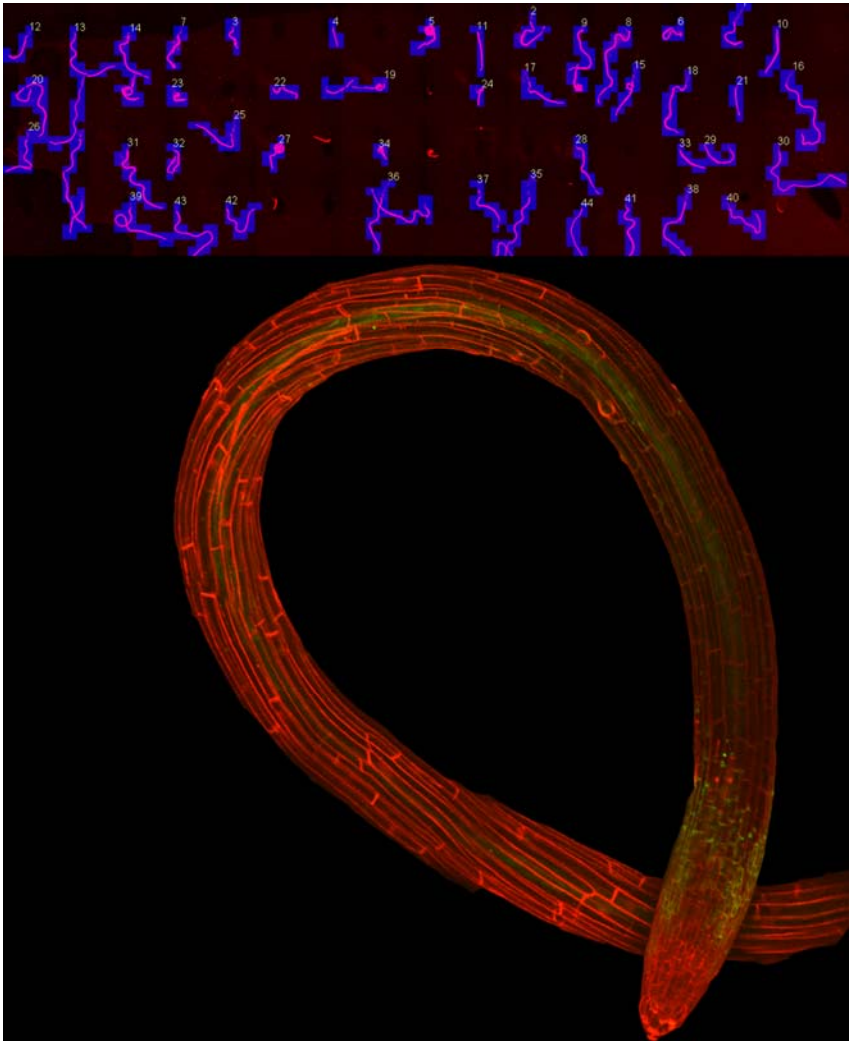
Abstracts of papers presented
at the 2010 meeting on

AUTOMATED IMAGING & HIGH-THROUGHPUT PHENOTYPING

December 5–December 8, 2010



Cold Spring Harbor Laboratory
Cold Spring Harbor, New York

Abstracts of papers presented
at the 2010 meeting on

AUTOMATED IMAGING & HIGH-THROUGHPUT PHENOTYPING

December 5–December 8, 2010

Arranged by

Philip Benfey, *Duke University*
Anne Carpenter, *Broad Institute of Harvard & MIT*
Robert Waterston, *University of Washington*
Uwe Ohler, *Duke University*

Cold Spring Harbor Laboratory
Cold Spring Harbor, New York

This meeting was funded in part by **LemnaTec GmbH;** and **Molecular Devices.**

Contributions from the following companies provide core support for the Cold Spring Harbor meetings program.

Corporate Sponsors

Agilent Technologies
AstraZeneca
BioVentures, Inc.
Bristol-Myers Squibb Company
Genentech, Inc.
GlaxoSmithKline
Life Technologies (Invitrogen & Applied Biosystems)
New England BioLabs, Inc.
OSI Pharmaceuticals, Inc.
Sanofi-Aventis

Plant Corporate Associates

Monsanto Company
Pioneer Hi-Bred International, Inc.

Foundations

Hudson-Alpha Institute for Biotechnology

Cover: High throughput imaging of Arabidopsis thaliana roots using the RootArray technology. Upper panel, automatic tracking of roots growing in the RootArray microfluidics device. Lower panel, maximum projection of a root image that was captured at high resolution using imaging coordinates that were determined by automated tracking. The image shows a root with stained cell walls (red) and a GFP reporter gene (green) indicating UPB1 mRNA expression. Figure by Wolfgang Busch.

AUTOMATED IMAGING AND HIGH-THROUGHPUT PHENOTYPING
Sunday, December 5 – Wednesday, December 8, 2010

Sunday	7:30 pm	**1** Cellular Phenotying
Monday	9:00 am	**2** Imaging Technologies and Platforms
Monday	2:00 pm	**3** Developmental Phenotyping
Monday	5:15 pm	**4** Future Directions
Monday	7:30 pm	**5** Poster Session and Wine & Cheese
Tuesday	9:00 am	**6** Stimulus / Response Phenotyping
Tuesday	2:00 pm	**7** Image Databasing and Knowledge Representation
Tuesday	4:45 pm	**8** Workshop on Databasing / Publication Guidelines
Tuesday	6:00 pm	Banquet
Wednesday	9:00 am	**9** Organismal Phenotyping

Mealtimes at Blackford Hall are as follows:
Breakfast 7:30 am-9:00 am
Lunch 11:30 am-1:30 pm
Dinner 5:30 pm-7:00 pm

Bar is open from 5:00 pm until late

Printed on 100% recycled paper.

PROGRAM

SUNDAY, December 5—7:30 PM

SESSION 1 CELLULAR PHENOTYPING

Chairperson: **R. Waterston,** University of Washington, Seattle

High content RNAi screens
Norbert Perrimon.
Presenter affiliation: Harvard Medical School, Boston, Massachusetts.

A phenomic assessment of sub-cellular morphology in *Saccharomyces cerevisiae* using automated genetics and high-content screening
Erin B. Styles, Lee Zamparo, Karen Founk, Mojca Mattiazzi, Jason Moffat, Zhaolei Zhang, Charles Boone, Brenda J. Andrews.
Presenter affiliation: Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada. 1

Quantifying phenotypes in physiologically relevant contexts
Anne E. Carpenter.
Presenter affiliation: Broad Institute of Harvard and MIT, Cambridge, Massachusetts. 2

Automated analysis of immunohistochemical images from the human protein atlas identifies potential location biomarkers related to cancer
Arvind Rao, Santhosh Bhavani, Estelle Glory, Robert F. Murphy.
Presenter affiliation: Carnegie Mellon University, Pittsburgh, Pennsylvania. 3

High-throughput mechanical cellular phenotyping by combined optical stretching and computational modeling
Evgeny Gladilin, Paula Gonzales, Josef A. Käs, Roland Eils.
Presenter affiliation: University of Heidelberg, Heidelberg, Germany; German Cancer Research Center, Heidelberg, Germany. 4

Lucas Pelkmans.
Presenter affiliation: Institute of Molecular Systems Biology, Zurich, Switzerland.

MONDAY, December 6—9:00 AM

SESSION 2 IMAGING TECHNOLOGIES AND PLATFORMS

Chairperson: **A. Carpenter,** Broad Institute of Harvard and MIT, Cambridge, Massachusetts

MONDAY, December 6—2:00 PM

SESSION 3 DEVELOPMENTAL PHENOTYPING

Chairperson: **F. Piano,** New York University, New York

MONDAY, December 6—5:15 PM

SESSION 4 FUTURE DIRECTIONS

MONDAY, December 6—7:30 PM

SESSION 5 POSTER SESSION and WINE & CHEESE PARTY

SESSION 6 STIMULUS / RESPONSE PHENOTYPING

Chairperson: **U. Ohler,** Duke University, Durham, North Carolina

TUESDAY, December 7—2:00 PM

SESSION 7 IMAGE DATABASING AND KNOWLEDGE REPRESENTATION

Chairperson: **P. Benfey,** Duke University, Durham, North Carolina

TUESDAY, December 7—4:45 PM

SESSION 8 WORKSHOP ON DATABASING / PUBLICATION GUIDELINES

TUESDAY, December 7

BANQUET

Cocktails 6:00 PM Dinner 6:45 PM

SESSION 9 ORGANISMAL PHENOTYPING

Chairperson: **S. Fraser,** California Institute of Technology, Pasadena

A PHENOMIC ASSESSMENT OF SUB-CELLULAR MORPHOLOGY IN SACCHAROMYCES CEREVISIAE USING AUTOMATED GENETICS AND HIGH-CONTENT SCREENING

Erin B Styles, Lee Zamparo, Karen Founk, Mojca Mattiazzi, Jason Moffat, Zhaolei Zhang, Charles Boone, Brenda J Andrews

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Molecular genetics, Toronto, M5S3E1, Canada

High-throughput (HTP) multi-channel fluorescence microscopy has enabled analysis of subcellular changes in the proteome of living cells in response to genetic and environmental perturbations. The budding yeast *S. cerevisiae* remains a premier model for the application of genome-wide approaches due to ongoing development of tools and reagents for systematic functional genomics and proteomics. In particular, collections of yeast strains are available with deletions of non-essential genes, conditional alleles of essential genes or carrying over-expression alleles of the entire genome. These collections, coupled with methods for automated yeast genetics (called Synthetic Genetic Array or SGA analysis), mean that genetic interactions can be rapidly assessed using a variety of phenotypic readouts, genome-wide.

We have undertaken a systematic, large-scale assessment of yeast sub-cellular morphology using both a comprehensive collection of fluorescent markers of cellular compartments and a series of sensitized genetic backgrounds. We coupled the SGA method and automated microscopy to quantitatively assess the abundance and localization of proteins in response to thousands of genetic perturbations. Our strategy involves using SGA to introduce fluorescent markers of key cellular compartments, along with sensitizing mutations, into the yeast deletion collection. Live cell imaging is then performed on the mutant arrays in early-log phase using HTP confocal microscopy. Our pipeline consists of quantitative image analysis to capture cellular morphology, texture and intensity measurements using CellProfiler™, followed by cellular classification using a support vector machine model to identify abnormal phenotypes. As proof-of-principle, we assessed DNA damage response and repair pathways by evaluating Rad52p-GFP foci in single mutants as well as in a number of chemically or genetically sensitized backgrounds (phleomycin treatment, *sgs1Δ*, *yku80Δ*). Computational analysis identified 93 mutants with increased levels of DNA damage foci, and analysis of mutants in sensitized backgrounds yielded an additional 152 mutants required for normal response to DNA damage. These results emphasize the importance of using sensitized genetic background to comprehensively identify new components of conserved biological pathways. We aim to expand our analysis to include many subcellular compartments, with the goal of producing a global view of subcellular morphology in a model eukaryotic system.

QUANTIFYING PHENOTYPES IN PHYSIOLOGICALLY RELEVANT CONTEXTS

Anne E Carpenter

Broad Institute of Harvard and MIT, Imaging Platform, Cambridge, MA, 02142

Image analysis is often straightforward when quantifying targeted phenotypes in simple monolayers of cultured cells. Our recent research focuses on algorithms for quantifying phenotypes in microscopy assays involving more physiologically relevant contexts. Using visually complex co-cultures of heterotypic cell types, for example, we are identifying inhibitors of leukemic stem cells and inducers of hepatocyte proliferation. As well, we are successfully screening the nematode *C. elegans* in microscopy assays that interrogate metabolism and infection in the context of a whole organism. Through novel assays and signature-extraction methods, we are also working to detect subtle morphological effects of disease-relevant perturbations.

AUTOMATED ANALYSIS OF IMMUNOHISTOCHEMICAL IMAGES FROM THE HUMAN PROTEIN ATLAS IDENTIFIES POTENTIAL LOCATION BIOMARKERS RELATED TO CANCER

Arvind Rao[1,2], Santhosh Bhavani[2], Estelle Glory[2,3], Robert F Murphy[1,2,3]

[1]Lane Center for Computational Biology, School of Computer Science, Pittsburgh, PA, 15213, [2]Center for Bioimage Informatics, Biomedical Engineering, Pittsburgh, PA, 15213, [3]Molecular and Bioimaging Center, Biological Sciences, Pittsburgh, PA, 15213

Significant efforts and progress have been made to identify candidate biomarkers that differ in expression between normal and cancerous tissues. Such biomarkers may be used for early detection, diagnosis, staging, prognosis, and patient-tailored therapy. In this study we seek to identify potential biomarkers that show differences in subcellular location instead of their quantitative expression. This approach automatically compares the subcellular location of proteins between normal and cancerous tissues in order to identify proteins whose location distribution is modified by oncogenesis. We have implemented an end-to-end workflow to identify such proteins using the tissue microarray collection composed of immunohistochemical (IHC) images of about 6000 proteins from the Human Protein Atlas [1, 2] (www.proteinatlas.org).

Methods: We extracted a number of subcellular location features from tissue images, such as texture descriptors from multi-resolution decomposition, and spatial co-localization information. The approach for the discovery of location biomarkers consists of two complementary analyses. In one analysis, nonparametric Friedman-Rafsky and k-nearest neighbor hypothesis tests are used to identify antibodies whose
image features change significantly between normal and cancerous tissue. Another analysis uses a classification method to identify proteins that have different subcellular location labels between normal and cancer conditions. Merging the lists from these two analyses makes the final list of potential biomarkers more robust to method biases.

Results: Based on antibody staining patterns for approximately 6000 proteins within thyroid, breast, pancreatic and prostate cancer, we identified an average of 7 potential location biomarkers in each tissue. Several of these proteins are known biomarkers in these four cancer types (based on changes in their expression levels). We conclude that the framework can provide lists of new location-biomarkers that are likely candidates for further exploration.

References:
[1] Uhlén M, Björling E, et.al (2005), A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics. 4(12):1920-32.
[2] Berglund L, Uhlén M, et.al (2008), A genecentric Human Protein Atlas for expression profiles based on antibodies, Mol Cell Proteomics. 7(10):2019--2027.

HIGH-THROUGHPUT MECHANICAL CELLULAR PHENOTYPING BY COMBINED OPTICAL STRETCHING AND COMPUTATIONAL MODELING

Evgeny Gladilin[1,2], Paula Gonzales[1,2], Josef A Käs[3], Roland Eils[1,2,4]

[1]University of Heidelberg, Bioquant, Heidelberg, 69120, Germany, [2]German Cancer Research Center, Theoretical Bioinformatics, Heidelberg, 69120, Germany, [3]University of Leipzig, Physics of Soft Matter, Leipzig, 04103, Germany, [4]University of Heidelberg, Institute of Pharmacy and Molecular Biotechnology, Heidelberg, 69120, Germany

Mechanical properties of the cell nucleus play an important role in maintaining the integrity of the genome and controlling the cellular force balance. The structural integrity of the nuclear interior is required for the simultaneous performance of essential biochemical processes such as replication, transcription and splicing. The nuclear functional architecture depends on the material properties of the cell nucleus. Irregularities in these properties have been related to a variety of force-dependent processes in the cell, such as migration, division, growth or differentiation. Characterizing the mechanical properties of the cell nucleus in situ and relating these parameters to cellular phenotypes or disease states remains a challenging task. Previous approaches of experimental cell mechanics are based on micromanipulation techniques that employ application of controlled forces onto the cellular boundary. Consequently, these methods provide information about the overall cell properties (e.g., stiffness of the entire cell) or those of its part that are directly accessible by the measurement (e.g., cell membrane). Probing mechanical properties of intra-cellular structures that are not accessible for direct measurement is not possible with conventional micromanipulation techniques. Furthermore, most conventional methods require time-consuming "one man – one cell" procedures restricting these approaches to very small numbers of experiments. Here, we present a general framework for large-scale functional mechanical cellular phenotyping that allows the determination of material properties for thousands of cells on a single cell basis. This approach combines contactless optical stretching of cells and model based analysis of time-series of microscopic images of these optically stretched cells. In a proof-of-concept study this framework was applied to estimate mechanical properties of various cell types under functional perturbations by RNAi and chemical drugs known to interfere with the cellular integrity. Combination of model-based image analysis with optical cell stretching paves the way for high-throughput molecular and mechanical cellular phenotyping with manifold applications in quantitative biology and molecular medicine.

SHEDDING LIGHT ON THE SYSTEM: RECONSTRUCTING DEVELOPMENT WITH LIGHT SHEET MICROSCOPY

Philipp J Keller[1], Annette D Schmidt[2], Anthony Santella[3], Khaled Khairy[1], Zhirong Bao[3], Jochen Wittbrodt[2,4], Ernst Stelzer[5,6]

[1]Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA, 20147, [2]University of Heidelberg, Department for Developmental Physiology, Heidelberg, 69120, Germany, [3]Sloan-Kettering Institute, Developmental Biology Program, New York, NY, 10065, [4]Karlsruhe Institute of Technology, Institute of Toxicology and Genetics, Karlsruhe, 76021, Germany, [5]European Molecular Biology Laboratory, Cell Biology and Biophysics Unit, Heidelberg, 69117, Germany, [6]Goethe University, Frankfurt Institute for Molecular Life Sciences, Frankfurt am Main, 60438, Germany

Embryonic development is one of the most complex processes encountered in biology. In vertebrates and higher invertebrates, a single cell is transformed into a fully functional organism comprising several tens of thousands cells, which are arranged in intricate organs and tissues able to perform the most impressive tasks. Although capturing and analyzing the morphogenetic dynamics of this process is crucial for basic research as well as for applied medical sciences, comprehensively reconstructing – and even recording – vertebrate embryogenesis has so far been impossible.

The novel light sheet-based microscopy technique DSLM allows recording the development of entire zebrafish and fruit fly embryos *in vivo* and with sub-cellular resolution[1]. By imaging at a speed of 1.5 billion volume elements per minute, data in the order of up to several tens of terabytes are acquired for each embryo over the time course of up to several days, i.e. up to stages, in which the embryo's major organs are in a functional state. By using automated image processing algorithms, the image data of each embryo is converted into a digital representation (the "digital embryo"), i.e. a database with comprehensive information about migratory tracks and divisions of the embryo's cells[2,3]. The digital embryos permit following single cells as a function of time and reveal the developmental blueprints of tissues and organs in the whole-embryo context.

Powerful synergies arise from combining the digital embryos with functional assays. Disease models and mutant phenotypes can now be analyzed and understood on a truly quantitative level. In the long-term perspective, I envision the digital embryos as a key to uncover the conserved and emerging rules of development.

[1] PJ Keller and EHK Stelzer, Current Opinion in Neurobiology, 18:624-32 (2008).
[2] PJ Keller *et al.*, Science, 322:1065-9 (2008).
[3] PJ Keller *et al.*, Nature Methods, 7:637-42 (2010).

QUANTIFYING FORM AND FUNCTION: HIGH-THROUGHPUT SCREENING FOR PLANT BIOTECHNOLOGY

Michael H Malone, Jasenka Benac, Hyundae Hong, Josh Rameaka, Daniel N Riggsbee, Keith A Koutsky

Monsanto Company, Biotechnology, Research Triangle Park, NC, 27709

Innovative technologies are required to meet the agricultural needs of a growing world population, which is estimated to reach 9 billion by 2050. Monsanto is committed to meeting these needs by improving the lives of farmers. Our goal is to use breeding, biotechnology, and improved agronomic practices to develop crops that produce more yield while conserving natural resources. A key step in this process lies in identifying plants that possess the traits that enable farmers to produce more yield with less water and fertilizer. To speed the identification of plants that have the traits farmers need, we have developed a high-throughput screening facility that fuses automated plant handling and imaging technology. This high-throughput screening facility allows us to quantify growth and whole plant physiology in precisely controlled environments. Here we present the opportunities and challenges inherent in high-throughput screening of biological systems, particularly as they apply to plant biotechnology.

TIME-RESOLVED PHENOTYPIC PROFILING FOR SECONDARY SCREENING BY TIME-LAPSE MICROSCOPY.

Thomas Walter[1], Moritz Mall[2], Michael Held[3], Matyas Gorjanacz[2], Iain Mattaj[2], Daniel Gerlich[3], Jan Ellenberg[1]

[1]EMBL, Cell Biology and Biophysics Unit, Heidelberg, 69117, Germany, [2]EMBL, Directors' Research, Heidelberg, 69117, Germany, [3]ETH, Institute of Biochemistry, Zurich, 8093, Switzerland

Over the last years, microscopy-based genetic and RNAi screens have become an indispensable tool to identify the genes required for various cellular processes. Recently, it was shown that even a highly dynamic process like the division of human cells can be studied on a genome wide scale by live cell imaging (Neumann, Walter et al., 2010; www.mitocheck.org). Such screens typically rely on computational pipelines to derive multidimensional phenotypic scores for cell populations allowing the clustering of genes into different phenotypic groups for which mechanistic predictions can be made.
The aim of secondary screening is then the detailed mechanistic analysis of the newly identified genes in these phenotypic groups and is typically carried out by imaging living cells at higher spatial and temporal resolution employing fluorescent reporters tailored to the mechanistic hypothesis. Secondary screening data therefore poses new challenges for image processing and requires analysis of dynamic changes at the single cell level. To address these computational challenges, we have developed CellCognition, a public domain software platform for the automatic processing of time-lapse microscopy data (Held et al., 2010; www.cellcognition.org). CellCognition features a growing portfolio of functional assays to unravel even subtle kinetic phenotypes of dynamic cellular processes. Here, we exploit this software to study genes with a presumed function in mitotic entry. The nuclear lamina, a meshwork of intermediate filaments that structurally supports the nuclear envelope, has to be disassembled in early mitosis to allow chromosome segregation. To study potential regulators of lamina disassembly, we imaged their knockdown or chemical inhibition phenotypes in HeLa cells stably expressing H2B-mCherry as a mitotic chromosome landmark and GFP-LaminB1 to monitor lamina depolymerization. Our approach combined supervised machine learning methods for the identification of the submitotic phases, automatic measurement of soluble and polymerized lamin and in silico alignment of single cell trajectories with time series analysis to quantify the timing of Lamin depolymerization and to relate it to the phase lengths in mitosis. With this powerful automatic analysis of single cell kinetics, we could reveal genes whose knockdown significantly delayed the lamina disassembly process.

Neumann, Walter et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature (2010) vol. 464 (7289) pp. 721-7
Held et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nat Methods (2010) vol. 7 (9) pp. 747-54

FLUORESCENCE CORRELATION SPECTROSCOPY AS TOOL FOR HIGH-CONTENT-SCREENING (HCS-FCS) IN *S. CEREVISIAE*

Christopher J Wood[1], Will A Marshall[1], Joseph Huff[1,3], Brian D Slaughter[2], Jay R Unruh[1,2], Winfried Wiegraebe[1]

[1]Stowers Institute for Medical Research, Stowers Microscopy Center, Kansas City, MO, 64110, [2]Stowers Institute for Medical Research, Research Advisors, Kansas City, MO, 64110, [3]Carl Zeiss, Microimaging, Thornwood, NY, 10594

To measure protein interactions, diffusion properties, and local concentrations in single cells, Fluorescence Correlation Spectroscopy (FCS) is a well established and widely accepted method. However, measurements are time-consuming and laborious. Therefore investigations are typically limited to ten, twenty or a few hundred cells. We developed an automated system to overcome these limitations and make FCS available for high-content-screening (HCS-FCS) in living cells.

In cellular auto-correlation FCS a focused laser-beam is positioned at a well-defined position within a cell. The protein of interest is tagged with a chromophore. Whenever one of these labeled proteins diffuses through the focal spot of the laser, the emitted photons are recorded. This data is evaluated using correlation functions and fitting of theoretical diffusion models to determine the average number of molecules and their residence time in the confocal volume. After calibration of this volume, we can calculate the local concentration and diffusion constant for the tagged protein.

An extension of auto-correlation FCS is cross-correlation FCS (FCCS). In this approach two proteins are labeled with two distinct dyes. Analysis of the fluctuation data using cross-correlation allows us finding interacting protein pairs.

We use confocal images and custom written software to detect single cells and to position the laser beam. The confocal images enable us to determine protein localization and local concentration for the whole cell.

We will present data from an auto-correlation screen of the eGFP tagged S. cerevisiae library developed at UCSF in the Weissman and O'Shea labs and distributed by Invitrogen. We acquired data for 4082 different proteins from more than 50.000 individual cells. The screen was performed in a 96 well plate format. This data allows us to assign global diffusion properties to individual cell compartments. The most important part of this investigation is the huge data-set of concentrations and diffusion constants for individual molecules in vivo that helps us to understand the dynamic behavior of these specific proteins in their cellular environment. Statistical analysis of this data-set gives us insight into common features of different cellular environments probed by these molecules.

For interaction screens we add a mCherry labeled copy of a protein-of-interest as bait to a sub-population of the eGFP library.

CELLULAR CHARACTERIZATION GOES HYPERSPECTRAL

Maria Cristina Pedroso[1], Howland D T Jones[2], Michael B Sinclair[2], David M Haaland[2]

[1]Monsanto Company, Crop Analytics & Automation, St. Louis, MO, 63167, [2]Sandia National Laboratories, Bioenergy and Defense Technologies, Albuquerque, NM, 87185-0895

Contrary to traditional confocal microscopes, the hyperspectral confocal fluorescence microscope has the ability to follow many spectrally and spatially overlapping tags simultaneously and can discriminate them against autofluorescence or impurity emissions. The tridimensional (3D) hyperspectral images obtained are then analyzed using Sandia's proprietary Multivariate Curve Resolution (MCR) algorithms. These algorithms can reveal all independent emitting components without a-priori information by extracting the emission spectrum of each component and providing relative concentration maps of each component for each voxel. This software is rapid and efficient to use and can handle large datasets in minutes instead of several hours. Additional software packages, specifically developed for batch processing and analysis of these plant image data, are being used in support of high throughput plant phenotype screening. Hyperspectral imaging of Arabidopsis thaliana and Zea mays (maize) mutants with various leaf phenotypes revealed significant changes in chloroplast pigment concentration, pigment ratios and spatial distribution per mutant leaf and chloroplast which correlated with wet chemistry and physiological data. Overall, this imaging system enables the observation and accurate quantification of plant fluorescent components and cellular processes never achieved before, therefore more in-depth knowledge of organelle development and microstructure can be seen with unprecedented resolution.

Acknowledgments

The support of Monsanto and Sandia management is gratefully acknowledged. Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

HIGH-THROUGHPUT IN VIVO CELLULAR-RESOLUTION WHOLE-ANIMAL SCREENING

Mehmet F Yanik

MIT, Engineering, Cambridge, MA 02139, MA, 02139

In recent years, the advantages of using small animals as models for human diseases have become increasingly apparent culminating several Nobel Prizes in Medicine and Chemistry. The availability of a wide array of genetic techniques, along with the animals' transparency, and ability to grow in minute volumes make the invertebrate C. elegans and the vertebrate zebrafish extremely versatile organisms. In particular, the highly complex and transparent organs of zebrafish allow sophisticated assays such as organ regeneration and neuronal function. Chemicals discovered using zebrafish such as FT1050 are already in clinical trials. However, over the last several decades, little has changed in how such multi-cellular organisms are manipulated, which has significantly limited both the scale and content of assays. We developed key technologies that allow cellular-resolution high-throughput manipulation and imaging of both C. elegans and zebrafish for large-scale genetic and drug discoveries (PNAS Aug. 2007, Nature Methods Aug. 2010). These technologies allow: (1) Loading of animals from multiwell plates compatible with industrial robotics, (2) Three-dimensional rotation and orientation of animals, (3) High-speed confocal imaging of entire organs at cellular resolution within few seconds, (4) Laser manipulations with subcellular precision such surgery, photolabeling and uncaging, (5) Dispensing and incubation of animals in chemical libraries. In combination with the femtosecond laser microsurgery technique we previously developed (Nature, Dec. 2004) for high-throughput neuronal injury; we performed in vivo neuronal regeneration screens on tens of thousands of animals, and identified compounds that enhance regeneration significantly following injury (PNAS, Oct 2010).

1. Rohde, C., Zeng, F., Gonzalez, R., Angel, M., Yanik, M. F., "Microfluidic system for on-chip high-throughput whole-animal sorting and screening at subcellular resolution", PNAS 104, 13891 (2007).
2. Pardo-Martin, C., Chang, T.-Y., Koo, B., Gilleland, C., Wasserman, S., Yanik, M. F., "High-throughput in vivo vertebrate screening", Nature Methods, July 19th (2010).
3. Commentary by Owen T. J. and Zon L. I., "Fishing at cellular level", Nature Methods, Aug 2010.
4. Yanik M. F. et al., "Neurosurgery: Functional Regeneration after Laser Axotomy," Nature 432, 822 (2004).
5. Samara, C., Rohde, C. B., Gilleland, C., Norton, S., Haggarty, S., Yanik, M. F., "Large-scale in vivo femtosecond laser neurosurgery screen reveals small-molecule enhancer of regeneration", PNAS, Oct. 2010.
6. Steinmeyer, J., Gilleland, C., Pardo, C., Angel, M., and M. F. Yanik, "Construction of a femtosecond laser microsurgery system", Nature Protocols 5, 395 (2010).
7. Gilleland C., Rohde C., Zeng F., Yanik M. F., "Microfluidic immobilization of physiologically active Caenorhabditis elegans", Nature Protocols, Nov 4th (2010).

A CELLULAR RESOLUTION ATLAS OF GENE EXPRESSION DYNAMICS IN *C. ELEGANS*

John I Murray[1,2], Thomas Boyle[2], Elicia Preston[1,2], Dionne Vafeados[2], Mihail Sarov[3], Robert Waterston[2]

[1]University of Pennsylvania, Department of Genetics, Philadelphia, PA, 19104, [2]University of Washington, Department of Genome Sciences, Seattle, WA, 98195, [3]Max Planck Institute of Molecular Cell Biology and Genetics, TransgeneOmics, Dresden, 01307, Germany

We recently developed methods to quantitatively measure gene expression dynamics with cellular resolution in *C. elegans* embryos. We have now used these methods to analyze the embryonic expression of over 150 genes, mostly transcription factors. The resulting cellular resolution atlas of gene expression dynamics is a powerful tool for studying embryonic fate specification.

We identified many genes whose expression was correlated with cell fate or, like Hox genes, position within the embryo. Surprisingly, we also found many genes whose expression was not well-correlated with these features but exhibited repetitive periodic lineage patterns. These different types of patterns combine to distinguish the two daughters of 65% of the divisions in early embryogenesis. We did not see many cell lineages with a single master fate regulator expressed specifically in only that lineage, however each lineage has an essentially unique expression pattern - we can distinguish over 99% of cell pairs in the embryo based on gene expression.

To look specifically for potential posttranscriptional regulation, we examined the expression dynamics of protein-GFP translational reporters. These reporters were largely expressed in the same cells as promoter-fusion reporters for the same gene, but often showed much more complicated dynamics. One common pattern was initial expression in a broad set of cells, followed by dynamic loss of expression in a subset of these that led to a more specific later pattern. Both the appearance and disappearance of TF expression was often tightly associated with cell division, and several genes were expressed specifically in all cells during a very limited developmental time window. Together, these results emphasize the complex regulation needed for a cell to adopt the correct fate at the correct time during development.

AUTONOMOUS SYNAPTOGENESIS SCREENING VIA SVM-GENERATED QUANTITATIVE PHENOTYPICAL SPACE

Matthew M Crane, Jeffrey N Stirman , Hang Lu

Georgia Tech, Bioengineering, Atlanta, GA, 30332

We present a comprehensive system composed of a computer vision framework using support vector machines (SVM) and an automated microfluidic system; we use it to phenotype and screen for synaptogenesis genes in C. elegans. Challenges in conventional approaches are as follows: (1) number of animals required for forward or reverse genetics is large (often 20,000+); (2) small feature size (synapses are often sub-micron and dimmer than autofluorescent background); (3) intensity, size, precise location, numbers, and co-localization are difficult to screen by eye; (4) manual manipulation and bias (fatigue) cause low throughput. Saturating most synaptic screens is currently not feasible because of the scale and labor required and the subtlety of the features.

In order to cope with the number of animals required for complete genome coverage in a forward screen, we have created a microfluidic system capable of processing 600-800 animals per hour. The device is capable of fully immobilizing animals for diffraction-limited imaging, and has been optimized to reduce the number of device failures to allow for uninterrupted imaging of >10,000 animals per device. The device is fully computer controlled, allowing for numerous error handling routines to enable continuous operation without an operator.

Computer vision provides signifanct advantages for autonomous high-throughput screens. Our algorithm is **multi-layered to ensure rapid, real-time identification, phenotyping and classification to allow screens continuously with minimal errors**. The reporter localizes on synapses of a motor neuron in the tail of C. elegans. Projections of z-stacks are used to compute local and region-specific features. Large-scale kernel SVM calculations are computationally expensive (resulting in minutes of processing). To allow on-line decision-making during a screen (~3s), fast elimination of locations unlikely to be synapses is performed using a linear SVM model; accuracy is ensured by an RBF-kernel SVM model subsequently applied to higher probability areas which are likely to contain the subcellular synapses. Once synapses are detected, a large number of synapse specific features (e.g. intensity, location, shape) are extracted and used to create the phenotypical space. Location on **this phenotypical space is used to determine if an animal is a novel mutant**.

The coupled microfluidic and computer vision system has allowed automated screening of thousands of animals, and **identified several classes of new mutants showing altered synapse numbers, locations and reporter intensity**. Our method enables image acquisition and quantitative phenotyping of synaptic expression a hundred fold faster than could be performed by hand – the only current alternative. This project will significantly affect the model organism community by transforming extremely challenging screens that would require months or years to complete, into simple, automated tasks that can be completed within a few weeks.

GLOBALLY OPTIMAL RECONSTRUCTION OF LARGE BIOLOGICAL SAMPLES IMAGED WITH HIGH-RESOLUTION MICROSCOPY TECHNIQUES.

Stephan Saalfeld, Stephan Preibisch, Pavel Tomancak

Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, D-01307, Germany

In the times when systems biology matures from buzzword of unclear meaning into increasingly popular scientific discipline it is important to provide techniques capable of monitoring the properties of the entire biological system. Microscopy is one of the premier approaches to gather information about spatial and temporal distribution of marked entities within biological systems. In particular in developmental biology, where the system means a tissue, organ or organism that can be very large, microscopy faces new challenges in achieving high-resolution across the entire imaged sample. We will discuss microscopy strategies commonly use to cover large biological systems with high-resolution such as 2d and 3d mosaicing and multi angle acquisition. We will present image analysis approaches that we developed to assemble the data into continuous representations of the large systems. As a common theme, the approaches use global optimization algorithms to prevent propagation of errors from consecutive registration steps. The microscopy techniques to which we adapted these strategies range from serial section electron microscopy of the brain tissue, through large, 3d confocal image mosaics, to multi-view recording of the entire living organisms with Selective Plane Illumination Microscopy (SPIM).

CELLULAR GENE EXPRESSION PROFILES IN DROSOPHILA BLASTODERM EMBRYOS

Zeba Wunderlich[1], Garth Ilsley[2], Meghan Bragdon[1], Kelly Eckenrode[1], Rolf Apweiler[2], Nick Luscombe[2], Angela DePace[1]

[1]Harvard Medical School, Dept. of Systems Biology, Boston, MA, 02130, [2]EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, United Kingdom

We employ semi-automated methods to capture cellular resolution gene expression profiles in fixed embryos of multiple Drosophila species. We use 2-photon microscopy to image thousands of individual embryos stained by fluorescent in situ hybridization or immunofluorescence, and process the resulting image stacks to build average models of RNA and protein expression for every cell over an hour of development prior to gastrulation. There is no cell division during this time, but thousands of genes are expressed at this stage, and many patterns are highly dynamic.

This data is particularly well-suited to computational approaches. We have used it to build models that uncover regulatory relationships between genes that pattern the blastoderm embryo. These models successfully predict known regulatory relationships for the well-characterized *even-skipped* locus. Importantly, they also predict the outcome of experimental perturbations. We are now scaling these models to predict regulatory relationships for uncharacterized genes, and to investigate alternative regulatory structures for the network.

We have also used this data to understand the origins of spatiotemporal differences in gene expression between closely related Drosophila species. Quantitative comparison of gene expression patterns between *Drosophila melanogaster, Drosophila yakuba* and *Drosophila pseudoobscura* has revealed multiple differences in their dynamics, relative position and proportion. We have used physically motivated modeling strategies to decipher the origins of these expression differences for a single gene, hunchback. Despite substantial *cis*-regulatory sequence differences, it appears that the species-specific differences in the expression of the hunchback posterior stripe are almost entirely explained by the differences in expression patterns of its regulators. The simple, interpretable nature of this modeling framework is easily extended to other genes and can be used to ask a variety of mechanistic questions about *cis*-regulatory module (CRM) function. These results also provide a framework to computationally screen for CRM sequence variants that have functional consequences, enabling us to rapidly curate CRMs sequences for further study.

TOWARDS HIGH THROUGHPUT SINGLE-CELL PHENOTYPING IN C. ELEGANS

Zhirong Bao, Zhuo Du, Julia Moore, Anthony Santella

Sloan-Kettering Institute, Developmental Biology Program, New York, NY, 10065

The invariant cell lineage of *C. elegans* has given single-cell resolution to its biology. Originated from the Waterston lab, we have developed an automated lineaging system Using 3D, time-lapse microscopy and image analysis, which tracks every cell at every minute through *C. elegans* embryogenesis and reconstruct the cell lineage, both in the wild type and in mutants. We are now further extending the lineaging system for high throughput and unbiased phenotyping. We have improved our imaging method so that we can image ~5,000 embryos a year on a spinning disc confocal. We have also developed new image analysis algorithms to reduce the error rate of automated lineaging so that one person can lineage ~1,000 embryos a year through all but the last round of cell divsion. In addition, we are developing statistical metrics to capture behaviors of each cell in the wild type and to systematically detect phenotypes in proliferation, differentiation and morphogenesis. We will present these recent developments as well as our initial effort to systematically phenotype genes required for embryogenesis.

MAPPING AND QUANTITATING CELLULAR PHENOTYPE, MORPHOLOGY AND GENE EXPRESSION THROUGHOUT *DROSOPHILA* EMBRYOGENESIS

David W Knowles[1], Soile V Keränen[1], Pablo Arbelaez[2], Jon T Barron[2], Mark D Biggin[1], Jitendra Malik[2]

[1]Lawrence Berkeley National Laboratory, Life Sciences and Genomics Divisions, Berkeley, CA, 94720, [2]University California at Berkeley, Electrical Engineering and Computer Sciences, Berkeley, CA, 94720

To quantitatively model animal transcription networks, it is essential to accurately measure the spatial and temporal expression patterns of transcription factors and their targets in their native 3D tissues structures. As part of the Berkeley *Drosophila* Transcription Network Project (BDTNP.lbl.gov), we have produced the first computationally analyzable, cellular resolution description of gene expression and morphology of the *Drosophila* blastoderm embryo. We are now extending this atlas through embryogenesis by developing automated, quantitative, high throughput image-based, learning and computer vision techniques.

Blastoderm embryos have a relatively simple structure, comprised of a single layer of 6000 cells surrounding a yolk. Over the course of the next ten hours and three mitotic cycles, however, large cell motions and complex patterns of differentiation lead to the formation of over 70 cell types and all the major larval organs.

To accurately capture this great increase in complexity, we have improved embryo imaging and are developing novel region specific nuclear and cellular segmentation strategies, establishing learning-based morphological classification methods to assign cells to specific tissues, and developing registration techniques to align multiple embryos images on a tissue by tissue basis.

Our strategy is to create fixed-cell atlases at multiple time points through embryogenesis, and to link these with cell-fate-maps created from live-cell images. The goal is to produce a computational-model, *VirtualEmbryo*, of *Drosophila* embryogenesis.

CHALLENGES FOR IMAGE EXPRESSION ANALYSIS: FROM AUTOMATED ACQUISITION TO HIGH-THROUGHPUT ANALYSIS

Bradley Martsberger, Bradley Moore, Iulian Pruteanu Malinici, Uwe Ohler

Duke University, Institute for Genome Sciences & Policy, Durham, NC, 27708

Advances in reporters for gene expression have made it possible to document and quantify expression patterns in 2D–4D. In contrast to microarrays, which provide data for many genes but averaged and/or at low resolution, images reveal the high spatial and/or temporal dynamics of gene expression. Developing computational methods to compare, annotate, and model gene expression based on images is imperative, considering that available data are rapidly increasing.

In this emerging field, the problems to solve are manifold and include: (a) Fully automated real-time acquisition of large numbers of biological samples; (b) extraction of meaningful information from large sets of potentially noisy images; and (c) models for spatial (expression) patterns extracted from image data. The challenge is to develop widely applicable methods that do not require large investments to adapt to a particular scenario. I will present ongoing work that addresses these questions in the context of biological model systems for development.

HIGH-THROUGHPUT DETERMINATION OF *C. ELEGANS* GROWTH AND VIABILITY PHENOTYPES

Erik C Andersen, Joshua S Bloom, Leonid Kruglyak

Princeton University, Lewis-Sigler Institute for Integrative Genomics, Princeton, NJ, 08544

Future successes in genetics depend on an increase in the throughput and accuracy of phenotypic measurements for two reasons. First, collections of strains with mutations in nearly every gene in the genome are available, so large numbers of independent strains need to be phenotyped. Second, we would like to know the phenotypic variance caused by genes as opposed to uncontrolled environmental and experimental factors. To more accurately quantify strain phenotypes, automation and assay replication help minimize the effect of experimental noise.

Using liquid propagation, synchronization, and scoring of *C. elegans* in 96-well plates, we have optimized quantitative growth assays. These assays allow for a single researcher to score nearly 200 strains a week. With quadruplicate assay replication and unbiased computer-assisted phenotyping, within and between assay variance is reduced. We measured a proxy of fecundity and developmental growth curves for over 100 wild strains to identify the most phenotypically diverse strains. In a parallel study, we used recombinant inbred lines derived from two genotypically and phenotypically diverse strains to map a quantitative traits locus for fecundity. We hope to expand these studies to identify phenotypic variation in susceptibilities to abiotic stresses and chemotherapy drugs in the wild isolates. Genome-wide association studies and linkage mapping will be used to identify quantitative trait loci that control these susceptibility traits. Additionally, we will present preliminary work on how these assays can be adapted to rapidly phenotype mutagenized populations allowing for identification of novel induced mutations.

IN VIVO QUANTIFICATION OF DYNAMIC GENE EXPRESSION IN THE *ARABIDOPSIS* ROOT

Wolfgang Busch[1], Bradley Martsberger[1], Bradley Moore[2], Richard W Twigg[1], Uwe Ohler[2], Philip N Benfey[1]

[1]Duke University, Department of Biology and IGSP Center for Systems Biology, Durham, NC, 27708, [2]Duke University, Institute for Genome Sciences and Policy, Durham, NC, 27708

Gene expression is a dynamic continuous phenomenon. To describe the expression of a gene it is therefore important to quantitatively capture spatio-temporal patterns of gene expression under defined conditions. As for many other quantitative experiments, it is essential to acquire measurements from multiple individuals. As yet, it is almost impossible to perform live imaging with cellular resolution on developing organs in multiple replicates and under different environmental conditions.
We have developed a microfluidics device, called the RootArray, which enables such studies. It permits more than 60 roots to be grown in parallel and to be imaged in short time intervals by confocal microscopy. The design of the RootArray allows for rapid exchange of growth media to alter environmental conditions and to observe subsequent alterations of gene expression. Our pipeline includes automated tracking and detection of growing roots, automated image acquisition and gene expression quantification. We systematically captured expression patterns of 12 genes in 24-48 h time courses. To assess dynamic properties of gene expression, we performed perturbations with changing media composition. Our dataset consists of high-resolution 3-dimensional images of more than 4000 roots to which more than 3,000,000 individual images contributed. We are currently automatically localizing and quantifying the GFP signals on those images, as well as determining growth characteristics of the roots.

MICROPILOT – AUTOMATION OF F-TECHNIQUES IN HCS

Christian Conrad[1], Annelie Wünsche[2], Tze Heng[2], Jutta Bulkescher[1], Rainer Pepperkok[1,2], Jan Ellenberg[2]

[1]EMBL, ALMF, Heidelberg, 69117, Germany, [2]EMBL, CBB, Heidelberg, 69117, Germany

Quantitative confocal fluorescence microscopy in cell biology, including advanced F-techniques (FRAP, FLIM, and FCS), relies on higher sample numbers to accomplish robust statistics. But those cell biological experiments are often hampered by tedious and subjective manual screening for defined cell characteristics, rare cellular morphologies, or specific events of interest. Especially in standardized High-Content Screening (HCS) of cell-based assays, it becomes increasingly more important to acquire massively single cell high-resolution image data in an unbiased manner. To efficiently identify such distinct cellular subpopulations or cell changes we engineered a 'Micropilot' microscope software suite facilitating supervised machine learning and the interface communication with various microscope systems during screening process. After labeling the targeted cell subpopulation in a representative training set, the software identifies the cell or cell event of interest automatically. Micropilot remotely communicates with different confocal/fluorescence microscopes to allow online reconfiguration of the microscopic devices and run advanced F-techniques protocols at defined locations. To prove this concept, mitotic phases (fixed and live) were automatically recognized during a fast low-resolution prescanning process and captured with higher spatial and temporal for further quantitative analysis. We anticipate basic communication to stop-and-continue on the screening acquisition side, as well as online-redefinition of the imaging protocols for HCS. In proof-of-concept experiments we could unravel novel protein properties by the sheer number of data points. Such automated platforms clearly provide a broad range of applications in cell and system biology.

GENE REGULATORY NETWORK OF CELL LINEAGE PROGRAM IN *C. ELEGANS* EMBRYOGENESIS

Zhuo Du, Anthony Santella, Julia Moore, Zhirong Bao

Sloan-Kettering Institute, Developmental Biology Program, New York, NY, 10065

The well characterized and invariant cell lineage makes *C. elegans* an excellent model organism for developmental biology. The invariant lineage implies that cell division, positioning and differentiation are tightly coordinated and regulated. However, the molecular mechanisms for establishing and executing a specific lineage program are not well understood. To analyze gene networks and pathways regulating cell lineage program we are performing systematic quantitative cell lineage analysis following genetic perturbations for specific genes. Through computational analysis of 3D time-lapse images, we automatically track every cell at every minute during embryogenesis and quantitatively measured cell behaviors such as cell division, positioning, cell cycle length and gene expression. We have identified ~200 high-penetrance, essential genes that potentially regulate various aspects of the embryonic cell lineage by large scale RNAi screen. For a decade, these genes have been known to be essential for embryogenesis but the underlying developmental mechanisms remain largely unknown. As a proof-of-concept study, 50 highly conserved genes were selected for in depth analysis. Currently, we have comprehensively analyzed ~100 cell lineages and associated single-cell phenotypes corresponding to loss-of-function of 25 genes that regulate specific aspects of the cell lineage program such as lineage-specific hyperplasia, lineage transformation and reprogramming. These preliminary data demonstrate that that our system is powerful for revealing mechanisms regulating the cell lineage program and provides a straightforward and efficient way to perform systematic analysis of *in vivo* gene function and genetic interaction networks.

PREDICTIVE CLUSTERING RELATES GENE ANNOTATIONS TO PHENOTYPE PROPERTIES EXTRACTED FROM IMAGES

Saso Dzeroski, Dragi Kocev

Jozef Stefan Institute, Dept. of Knowledge Technologies, Ljubljana, SI-1000, Slovenia

We address the task of grouping genes resulting in highly similar phenotypes upon siRNA mediated
downregulation. The phenotypes are described by features extracted from images of the corresponding cellular assays. Both freely available general-purpose software, such as CellProfiler, and custom-made proprietary software can be used for this purpose. The features capture properties (such as intensity or texture) of the cells or their parts (nuclei, citoplasm, Golgi apparatus ...) in the images.

Clustering produces partitions of the objects of interest (genes) into groups that are similar in a given feature space. In the context of the application of interest, this is a set of features extracted from the images of cellular assays. Besides finding clusters, e.g., groups of genes, we also aim to find descriptions/ explanations for the clusters. The groups are explained in terms of a set of descriptors from a separate space, i.e.,annotations of genes in terms of, e.g., the Gene Ontology or the KEGG Pathway Database.

The typical approach to the problem at hand is to first clusteri the phenotypes and elucidate the characteristics of the obtained clusters later on. Instead, we perform so-called constrained clustering, which
yields both the clusters and their symbolic descriptions all in one step. The constrained clustering is performed by using predictive clustering trees (PCTs): These exemplify the paradigm of predictive clustering, which combines clustering and prediction.

In the presentation, we will describe the method of building predictive clustering trees. We will also describe its application to the analysis of image data resulting from siRNA screens. The approach has been used to analyze image data from a siRNA screen designed to study MHC Class II antigen presentation.

As a result of the predictive clustering process, we obtain clearly defined/described groups of genes,
which yield highly similar phenotypes upon siRNA mediated downregulation. For example, one such group are the genes involved in the biological processes of 'defense response' and 'regulation of metabolic processes'. Groups of this kind can be used to identify pathways regulating the processes of interest (such as MHC Class II antigen presentation).

INTRA- AND INTER-GENOTYPE ROOT INTERACTION IN RICE AND ITS ANALYSIS BASED ON 3D RECONSTRUCTION

Suqin Fang[1,2], Randy T Clark[3], Joshua S Weitz[4], Hong Liao[2], Philip N Benfey[1]

[1]Duke University, Department of Biology and Institute for Genome Science and Policy Center for Systems Biology, Durham, NC, 27705, [2]South China Agricultural University, Root Biology Center, Guangzhou, 510642, China, [3]Cornell University, Department of Biological and Environmental Engineering, Ithaca, NY, 14853, [4]Georgia Institute of Technology, School of Biology and Physics, Atlanta, GA, 30332

Root systems play important roles in nutrients and water absorption as well as in anchorage. Plants have the ability to adjust their root systems to adapt to heterogeneous environments for better growth. We know that neighboring plants can cause dramatic phenotypic responses in many plant species. However, little is known about how a plant should respond to intra-genotype or inter-genotype root interaction. In most plants, root systems are below-ground structures which cause difficulty in direct observation of root interactions among neighboring plants. In this study, rice plants were grown in the transparent gel system which allowed capturing root system images nondestructively by a camera imaging platform. Two rice plants of the same genotypes or different genotypes, with 2 cm interval, were grown in one cylinder to study the intra-genotype or inter-genotype root interaction respectively. Imaging results showed that root systems of the same genotype plants grew toward each other while those of different genotype plants had the trend of growing away from each other. In shoot-separated experiments, where the two rice shoot systems were grown in two separate compartments to remove shoot interaction between the two plants, root systems showed the same growth trends as in shoot non-separated experiments respectively. To quantify the root interaction, 3D rice root systems were reconstructed based on a set of root images taken from different angles and the voxels of each rice root system were obtained, from which the root intersection was calculated. Analysis results showed that the root intersection of intra-genotype rice plants was significantly higher than that of inter-genotype in both shoot non-separated experiments and shoot-separated experiments, which indicated that plant root systems could not only sense but also identify the neighboring plants.

MAPPING GENE FUNCTION THROUGH IMAGE-BASED SYNTHETIC GENETIC INTERACTION ANALYSIS BY RNAI

Bernd Fischer[1], Thomas Horn[2], Thomas Sandmann[2], Michael Boutros[2], Wolfgang Huber[1]

[1]EMBL, Genome Biology, Heidelberg, 69117, Germany, [2]DKFZ, Division Signaling and Functional Genomics, Heidelberg, 69120, Germany

Large-scale mapping of genetic interactions has provided important new insights into the functional organisation of the yeast cell. However, the data show limited specificity due to the mapping of a broad 'cell growth' phenotype. Data for higher organisms has been missing.

We developed a computational method to quantitatively assess imaging-based phenotypes from pairwise RNAi-perturbations by automated image analysis, statistical processing and machine learning. In contrast to current data, which only provide single valued readouts, we measured multi-dimensional phenotype, which provide a more specific discrimination of the effects of different treatments on cellular processes. I will describe the statistical modeling that turns these raw phenotypic measurements into a quantitative matrix of genetic interactions.
From that, we used unsupervised learning for unbiased exploration of the data, and supervised learning for extrapolation of known signal transduction networks to unknown components and their interactions. We predicted new functions for previously uncharacterized genes, including a novel regulator of Ras/MAPK signaling, whose conserved function could be validated experimentally in vitro and in vivo.

In contrast to full knock-outs, RNAi produces hypomorphic phenotypes, and thus may provide a better model for the exploration of the effects of allelic series. However, the individual genetic interactions can be subtle, with presence or absence, as well as direction potentially depending on siRNA concentations and genetic background. Genetic interaction profiles, that is, the vector of all interactions of a gene with everyone else, by way of similarity-based (bi)-clustering, provide a robust and information-rich resource for functional discovery.

INTEGRATIVE IMAGE ANALYSIS OF DROSOPHILA IN SITU HYBRIDIZATION DATA

Charlie Frogner[1], Chris Bristow[2], Tom Morgan[2], Stan Nikolov[2], Anna Ayuso[2], Tomaso Poggio[1,2], Manolis Kellis[2]

[1]MIT, Brain and Cognitive Sciences, Cambridge, MA, 02139, [2]MIT, CSAIL, Cambridge, MA, 02139

Understanding the spatio-temporal control of gene expression during development is one of the major challenges in developmental biology, but image processing and genome analysis have traditionally remained separate. To address this challenge, we have sought to combine image analysis of gene expression patterns with sequence analysis of genomic data to explore the regulation of mRNA expression patterns during Drosophila embryogenesis. We have applied these methods to large-scale in situ hybridization screens that map spatial patterns of expression for thousands of genes across several developmental stages, seeking to dissect the regulatory constructs that control tissue-specific gene expression.

We used ~75,000 images of Drosophila embryos, encompassing ~6,000 genes profiled at 5 stage ranges during embryogenesis. We developed an image processing pipeline that registers embryos across different images and extracts the pixels that correspond to mRNA staining. This pipeline incorporates a novel approach for embryo segmentation that reliably addresses widespread challenges, such as multiple overlapping embryos. In addition, we introduced a new method for extracting detailed stain patterns from the images using a supervised learning approach trained on ubiquitously-expressed genes. Overall, our methods allowed us to extract mRNA expression patterns for over 68,000 embryos, with minimal human inspection.

To characterize the accuracy of each stage of our image processing pipeline, we developed a protocol that uses images hand-labeled with contours for evaluating segmentation as well as correspondence points between embryos for evaluating embryo registration. To acquire a large corpus of manually-labeled images, we coupled the open source LabelMe image annotation tool with Amazon's Mechanical Turk. Motivated by the importance of human annotations more generally in analyzing large image datasets, we are developing a significantly-extended version of LabelMe, called LabelLife. LabelLife will enable rapid annotation of images, parts of images and whole image sets with terms taken from controlled vocabularies, and will incorporate machine learning to predict annotations where possible.

To validate our image processing approach, we demonstrated correlations between transcription factor binding data and the extracted gene expression patterns. Following on this result, we are actively pursuing the integration of the image data with multiple genomics datasets, to uncover the regulatory programs governing spatio-temporal patterning of gene expression during Drosophila embryogenesis. We expect that our approach will transfer to many organisms, offering new insights into tissue-specific gene expression.

GIA-ROOTS: SOFTWARE FOR THE HIGH THROUGHPUT ANALYSIS OF PLANT ROOT SYSTEM IMAGES

Taras Galkovskyi[1], Yuriy Mileyko[2], Olga Symonova[3], Charles A Price[3], Joshua S Weitz[3,4], John Harer[2]

[1]Duke University, Biology Department, Durham, NC, 27708, [2]Duke University, Mathematics Department, Durham, NC, 27708, [3]Georgia Institute of Technology, School of Biology, Atlanta, GA, 30332, [4]Georgia Institute of Technology, School of Physics, Atlanta, GA, 30332

Breeders and scientists have begun to screen plant root systems in a high-throughput fashion as part of efforts to modify and improve crop function. The development of platforms for imaging plant root systems has led to a growing dilemma: how to analyze thousands, if not tens or hundreds of thousands, of root system images in an efficient and reproducible fashion. The dilemma is made more difficult because analysis of biological networks, including root networks, presents distinct computational challenges when compared to many well-supported software tools used in the analysis of cells and cell-like images. As such, we have developed a software tool, GiA-Roots (General Image Analysis of Roots), which automates the process of image processing and feature extraction of plant root systems, and packages its capabilities in a easy-to-use GUI framework. Specifically, GiA-Roots efficiently estimates quantitative network features from 2D and 3D images of plant root networks, involves the user in the analysis process without the burden of learning technical details of processing, and keeps track of all intermediate stages in image processing for full reproducibility of feature extraction. These objectives are accomplished by combining an intuitive graphical front-end for all users with an extensible and robust software architecture for technical users. We describe the GiA-Roots software focusing on the extensibility of the set of features that it can compute, thus providing a framework for many different extensions in the area of analyzing biological networks. As one example, we describe the Topological Root Analysis eXtension (TRAX) that computes a set of novel topological features capable of capturing subtle differences in root (and possible other kind of) network images.

IMAGE PROCESSING TOOLS FOR DEVELOPING FUNCTION SPACE MODELS OF LOCAL AND GLOBAL GROWTH PATTERNS IN ACTIVELY DEVELOPING STEM CELL NICHES

Venugopala G Reddy[1], Anirban Chakraborty[2], Moses Tataw[2], Ramkishor Yadav[1], Amit K Roy Chowdhury[2]

[1]Institute of Integrative Genome Boiology, Botany and Plant Sciences, Riverside, CA, 92521, [2]Center for Research in Intelligent Systems, Electrical Engineering, Riverside, CA, 92521

Pattern formation in developmental fields involves precise spatial arrangement of different cell types in a dynamic landscape wherein cells exhibit a variety of behaviors such as cell division, cell expansion and cell migration. Therefore, a quantitative understanding of the spatio-temporal dynamics of gene expression and cell behavioral patterns is crucial to model the process of development. We have developed live-imaging platforms to record gene expression and cell behaviors from actively developing shoot apical meristem (SAM)-stem-cell niches of Arabidopsis. SAM is a multilayered structure (about 500 cells) with stem cells located at the tip which are surrounded by differentiating daughters that develop into organ primordia in a specific spatio-temporal sequence. We are developing an image analysis pipeline that will be able to automatically analyze the growth characteristics of both individual cells and also groups of cells (organ primordia). Towards this goal, we have a number of challenges that are being overcome using a variety of tools from image processing, machine learning and statistical modeling. We have developed graph theoretic methods for computing cell lineages of hundreds of cells over 48-72 hours of time-lapse imagery. We are developing estimation theoretic methods for computing the cell volume from a sparse cell of cross-sectional slices. Combined with the lineages, this will allow us to learn various statistics of growth, including cell volume change, division rates, and mean time between divisions. A computational tool-kit will enable us to repeat these for various plants in which cell-cell communication is transiently-impaired. At a global level, we are developing methods for computing the growth rates of organ primordia by quantifying the changes in their shape as a function of time. The growth curves of primordia from different SAMs can then be aligned to understand the spatio-temporal dynamics of hundreds of genes expressed in distinct domains of the stem cell niche, leading to the development of a dynamic gene expression atlas.

EXTRACTING MICROSCOPY-BASED SIGNATURES OF HISTONE DEACETYLASE FUNCTION USING QUANTITATIVE IMAGE ANALYSIS

Sigrun Gustafsdottir*[1], Melissa Kemp*[1], Katherine Madden[2], Vebjorn Ljosa[2], Joshiawa Paulk[1], Candice Thompson[1], Deepika Walpita[1], J.Anthony Wilson[1], Paul Clemons[1], Anne Carpenter[2], Angela Koehler[1], Alykhan Shamji[1]

[1]Broad Institute of MIT and Harvard, Chemical Biology Program, Cambridge, MA, 02142, [2]Broad Institute of MIT and Harvard, Imaging Platform, Cambridge, MA, 02142

*Authors contributed equally

In conventional high-throughput image-based screens, fluorescent antibody conjugates or fluorescent protein fusions are typically used to determine the concentration and localization of a single target in the cell. The assay optimization is often time consuming and the resulting assay measures only a narrow range of biological outcomes. Our goal was to "paint" cells with as many fluorescent morphological markers as possible to capture greater information content about cellular morphology. We have generated a protocol that allows detection of seven cell compartments to report on hundreds of cellular features. Combined with advances in automated image analysis for extraction of quantitative information (CellProfiler), the assay has the potential to reveal morphological changes that are unexpected or even invisible to the human eye.

One area of interest has been to use this method to explore the biological functions and perturbations of histone deacetylase (HDACs). The nine HDACs in humans are thought to play a variety of roles, especially in transcriptional regulation. HDACs have been linked to a variety of cancers, psychiatric diseases, and other disorders, and inhibitors of these enzymes are currently in clinical trials. Characterizing the functional differences among the nine human HDACs, and identifying suitable chemical regulators for them, is an important frontier with therapeutic relevance. We are therefore using the multiplexed microscopy assay to identify image-based signatures and chemical inhibitors for each HDAC, using RNAi knockdowns and small molecule inhibitors. If successful, this proof-of-concept work could lead to applying the cell-painting/signature extraction technique to provide novel insights into chemically induced states, reveal mechanisms of action for small molecules, aid in the identification of cellular disease signatures, and ultimately enable the discovery of compounds that can reverse the diseased states.

PHENOTYPIC PROFILING OF BIOACTIVE LIBRARIES USING PRIMARY NEURONS

Omar Gutierrez-Arenas[1], Hassan Al-Ali[1], Stephan Schuerer[3], John L Bixby[1,2], Vance P Lemmon[1]

[1]University of Miami, The Miami Project to Cure Paralysis, Miami, FL, 33136, [2]University of Miami, Dept. of Molecular and Cell. Pharmacology, Miami, FL, 33136, [3]University of Miami, Center for Computational Sciences, Miami, FL, 33136

The high hit rates and functional diversities of small bioactive libraries like LOPAC, Spectrum, and EMD Inhibitor Select provide not only a source of lead compounds or candidates for drug "repurposing", but also an excellent set of profiling tools. The activity profile of the library on a biological substrate, such as a specific type of neuron in a specific environment, can be used both to characterize the biological object and to assign functional annotations to individual compounds. That is, the screen serves both to implicate known bioactivities in the production of cellular phenotypes, and to characterize particular compounds by the phenotypes they produce. In this work we have used a variety of primary neuronal culture systems to screen a library of 240 protein kinase inhibitors (EMD Inhibitor Select) using High Content Analysis. The cultures we tested differed in cell type (E18 hippocampal neurons, P3 cortical neurons, cerebellar granule cells, and dorsal root ganglion cells), coating substrate (polylysine, laminin, chondroitin sulfate proteoglycans) and media (Neurobasal/B27 with and without glia conditioning). Each compound was tested at five concentrations in each assay and each assay was performed twice; cell survival and neuronal morphology traits were measured. The activity profiles of the individual compounds varied significantly among the culture conditions. Some compounds ("hits") were identified as promoting neurite growth in one or more conditions; these are of interest for therapeutic strategies. Other compounds were toxic, and the overall toxicity profile of the library was decreased by glial conditioned medium. We built digital phenotypic vectors for each compound by concatenating values of 1, -1, or 0 (increasing, decreasing, or no change) for each phenotypic variable measured, and computed pairwise similarities using Tanimoto rules. Interestingly, some compounds with similar annotated specificity profiles against protein kinases produced substantially different phenotypic vectors. Our study highlights the usefulness of phenotypic profiling to elucidate compound activities beyond those obtained from binding profiles or enzyme inhibition panels.

Supported by NIH (HD057632), U.S. Army (W81XWH-05-1-0061), Paralyzed Veterans of America, and the Buoniconti Fund. VPL holds the Walter G. Ross Distinguished Chair in Developmental Neuroscience.

COMPARISON OF CLASSIFICATION STRATEGIES FOR THE AUTOMATED ASSESSMENT OF STEM CELL COLONIES

Michael Halter[1], Daniel J Hoeppner[2], John T Elliott[1], Ronald D McKay[3], Anne L Plant[1]

[1]NIST, Biochemical Science Division, Gaithersburg, MD, 20879, [2]NIH, Laboratory of Molecular Biology, NINDS, Bethesda, MD, 20892, [3]Lieber Institute, Lieber Institute for Brain Research, Baltimore, MD, 21205

The correct identification and selection of desirable stem cell colonies is a critical step in the practical use of stem cells for research and in the expansion of cultures for therapy. Methods such as antibody staining perturb the culture and are not ideal for use in routine cell maintenance. If visual inspection by transmitted light microscopy can be used robustly to select desired colonies, this would be a great advantage to the culture and expansion of stem cells. Experts that are highly trained in examining colonies can distinguish colonies of cells that are undifferentiated and pluripotent from colonies that have less desirable characteristics. If these differences can be quantified, this may provide a more complete understanding of the differences between these cells, and importantly, will make it possible for less trained biologists to correctly distinguish ideal from non-ideal colonies. In an examination of automated approaches for classifying human embryonic stem cell colonies, we compared the performance of classification schemes where features were calculated over the whole image, where features were calculated on image tiles, and a third approach where a 'bag of words' model was used. All three approaches use Haralick texture and wavelet features. In addition to developing an automated classification approach for stem cell colonies, this study is helping to define a quantitative description of image characteristics that are present in, and characterize, undifferentiated stem cell colonies.

QUANTIFICATION OF DYNAMIC MORPHOLOGICAL PROCESSES AND DRUG RESPONSES IN MINIATURIZED 3D PROSTATE CANCER CULTURES BY AUTOMATED IMAGE ANALYSIS

Ville Härmä[1], Antti Happonen*[2], Johannes Virtanen[3], Jyrki Lötjönen[4], Harri Siitari[1], Matthias Nees[1]

[1]VTT Technical Research Centre, Medical Biotechnology, Turku, 20521, Finland, [2]University of Technology, Signal Processing, Tampere, 33101, Finland, [3]University of Turku, Biotechnology Centre, Turku, 20521, Finland, [4]VTT Technical Research Centre, Knowledge Intensive Services, Tampere, 33101, Finland

Prostate cancer (PrCa) cells embedded in three-dimensional (3D) organotypic cell culture, form variable epithelial morphologies. These phenotypes result from differences in epithelial differentiation patterns compromised in cancer vs. normal cells, and reflect corresponding differences between normal and transformed human tissues *in vivo*. The degree of epithelial maturation, cell-cell contacts, cell shape and polarization are measures for residual differentiation capacity in cancers, and are useful as sensitive and dynamic parameters to evaluate the response of PrCa to changes in the microenvironment, altered growth conditions, or chemotherapeutic compounds.

We have developed 3D cell culture model systems for the screening of functional gene knock-down and small molecule inhibitors. Miniaturization facilitates robust and reproducible high-content screens in 3-D platforms. We have further designed a panel of automated image analysis software tools to enable the real-time quantitative analysis of morphological changes in 3D, based on live-cell microscopic imaging. This is complemented by statistical post-processing of morphological image data as an integrated readout system for biomedical drug discovery applications.

The most relevant cancer phenotype, reflecting the aggressive, invasive nature of advanced PrCa's, is the formation of "stellate" structures that rapidly penetrate the surrounding ECM. Morphological transformation of organotypic spheroids may also occur spontaneously, indicating an extreme epithelial plasticity in some PrCa cell lines. This corresponds to intrinsic transdifferentiation programs such as epithelial-to-mesenchymal transition (EMT), considered to play a key role in cancer cell invasion and metastasis. Spheroids undergoing spontaneous EMT are rapidly transformed from normal-like, round spheroids into stellate structures, and represent a unique model to investigate molecular mechanisms and decision-making, relevant for invasiveness and motility. Drug and siRNA screens have identified G-protein coupled receptors and lipid metabolites as key mechanisms regulating cell motility and EMT. The bioinformatic analysis of underlying gene networks in large-scale PrCa transcriptome studies validates the relevance of such mechanisms also in clinical cases and metastases.

AUTOMATED PHENOTYPING FOR QTL ANALYSIS OF RICE ROOT SYSTEM ARCHITECTURE

Anjali S Iyer-Pascuzzi[1,2], Christopher N Topp[1,2], Jill T Anderson[1], Ying Zheng[3], Yuriy Mileyko[4], John L Harer[4,2], Herbert Edelsbrunner[3], Joshua S Weitz[5,6], Thomas Mitchell-Olds[1], Philip N Benfey[1,2]

[1]Duke University, Biology , Durham, NC, 27708, [2]Duke University, IGSP Center for Systems Biology, Durham, NC, 27708, [3]Duke University, Computer Science, Durham, NC, 27708, [4]Duke University, Mathematics, Durham, NC, 27708, [5]Georgia Institute of Technology, Biology, Atlanta, GA, 30332, [6]Georgia Institute of Technology, Physics, Atlanta, GA, 30332

Plant health and survival is dependent on the root system architecture (RSA), the spatial configuration of different types and ages of roots on a single plant. Root systems are highly plastic, allowing for soil exploration in diverse conditions. Modification of RSA could contribute to improvements of desirable agronomic traits such as drought tolerance and resistance to nutrient deficiencies. Although roots are central to plant fitness, knowledge regarding the genes underlying RSA is limited, in part due to the inaccessibility of root systems. We have recently developed a non-destructive gel-based imaging and analysis system for automated phenotyping of root system architecture in three dimensions. Here, we use this system for QTL analysis of rice root architecture. We imaged and automatically phenotyped 16 traits in the root systems of 180 recombinant inbred lines of rice under nutrient replete conditions across three days. We find multiple QTL on each day, several of which correspond to those previously identified using sand or soil-based systems. In addition, we explore the effects of different nutrient deficiencies on the root system. This work forms the foundation for fine-mapping and cloning the genes responsible for root system architecture in a variety of environmental conditions.

CELL LINEAGE TRACKING TO IDENTIFY AND REVEAL THE STRATEGIES FOR CLONAL EXPANSION OF A TUMOUR SYSTEM - INFLUENCES AND RESPONSES TO TOPOISOMERASE I INHIBITION

Imtiaz A Khan[1,2], Ricardo Santiago-mozos[5], Nick S White[3], Janet Fisher[2], Richard J White[4], Michael Madden[5], Paul J Smith[2], Rachel J Errington[2]

[1]Broad Institute of MIT and Harvard, Imaging Platform, Cambridge, MA, 02142, [2]Cardiff University, School of Medicine, Cardiff, CF14 4XN, United Kingdom, [3]Cardiff University, School of Optometry and Vision Sciences, Cardiff, CF24 4LU, United Kingdom, [4]Cardiff University, School of Computer Sciences, Cardiff, CF24 3AA, United Kingdom, [5]National University of Ireland, Information Technologies, Galway, Ireland

Strategies for responding and overcoming perturbation that leads towards the accrual of drug resistance in the treatment of cancer are constant themes in drug discovery and cancer research. The efficacy of anticancer drug, topotecan has shown to impart its acute effects on cells, but long term clonogenic experiment showed the appearance of cell colonies generated from resistant subpopulation of cells. The strategies undertaken by these subpopulation of cells to maximize clonal expansion are still unknown. Here exploiting timelapse derived images through bioinformatics and machine learning techniques, we have dissected the impact and operation of asymmetric events and interrelationships of cells and thus revealed the strategies for clonal expansion. By mapping pharmacodynamic response on a bifurcating lineage structure, we found that cells adopt a stem cell like differential behaviour, both at spatial and temporal domain. Such structure reiterated the importance of cell cycle positioning of the progenitor generation, as the key factor for survival potential, and at the same time identified the critical bifurcating points in successive generation that can illustrate the strategies for clonal expansion. This multidisciplinary approach of mapping dynamic cellular behaviour on bifurcating lineage map facilitated us to transform images to knowledge. Through such transformation we have revealed the 'informative cells' underpinning the clonogenic expansion potential of a perturbed cellular systems. We anticipate that understanding the epigenetic structure of these informative cells would enhance the efficacy of cancer therapy.

CYCLoPs: A COMPREHENSIVE DATABASE OF YEAST CELL IMAGES, SUB-CELLULAR LOCALIZATION AND PROTEIN ABUNDANCE FOLLOWING CHEMICAL AND GENETIC PERTURBATION

Judice L Koh[1], Yolanda Chong[1], Alan Moses[2], Brenda J Andrews[1], Jason Moffat[1]

[1]University of Toronto, Banting and Best Department of Medical Research. The Terrence Donnelly Centre for Cellular and Bio-molecular Research, Toronto, M5S 3E1, Canada, [2]University of Toronto, Department of Cells and System Biology, Toronto, M5S 3B2, Canada

Eukaryotic cells are organized into sub-cellular compartments that serve as hubs for all molecular interactions within a cell. Under different environmental cues, groups of proteins are shuttled between sub-cellular localization sites to carry out specific cellular functions. Descriptive mapping of the localization dynamics or flux within the cell provides an invaluable resource toward elucidating specific mechanisms and pathways within the cell.

Using a combinatorial approach of High-Content Screening, large-scale Synthetic Genetic Array (SGA) technique and machine learning algorithms for characterizing morphological patterns from cell images, we obtained quantitative descriptions of the localization dynamics within the proteome of *Saccharomyces cerevisiae* in various genetic and chemical backgrounds. This approach has generated an invaluable resource towards elucidating specific mechanisms and pathways in the cell and has produced a collection of more than 600,000 micrographs, comprising 40 million cells and ~400,000 quantitative measurements depicting the temporal localization dynamics as well as the protein levels of ~4,500 genes under 4 chemical treatments and 11 genetic knockouts. The collection is available to the community through **CYCLoPs** (**C**ollection of **Y**east **C**ells **Lo**calization **P**attern**s**) - a web-based image database that aims at providing a central platform for searching, visualizing and analyzing yeast cells and for querying the localization and abundance measurements of the yeast proteome.

CYCLoPs utilizes mySQL as the relational database management system, with graphical interface developed in Java Script, R, CGI-Perl, SQL and HTML. It allows users to query the localization and abundance profiles of a specified gene, and retrieve top genes that are transported between one or more of the 16 sub-cellular compartments or exhibit changes in protein levels in the presence of drug treatment or gene deletion. Search results and images may be downloaded.

OVEREXPRESSION OF PERIPHERAL NERVOUS SYSTEM GENES IN CENTRAL NEURONS INDUCES GROWTH ON INHIBITORY SUBSTRATES

Vance P Lemmon[1,2], Buchser J Buchser[1], Robin P Smith[1], Jose R Pardinas[1], Candace L Haddox[1], Stanley R Hoffman[4], John L Bixby[1,3]

[1]Univ. of Miami, Miami Project to Cure Paralysis, Miami, FL, 33136, [2]Univ. of Miami, Neurological Surgery, Miami, FL, 33136, [3]Univ. of Miami, Mol & Cell Pharmacology, Miami, FL, 33136, [4]Medical Univ. of South Carolina, Rheumatology, Charleston, SC, 29425

Trauma to the central nervous system (CNS) can result in irreparable loss of function, owing to the failure of axonal regeneration from CNS neurons. This failure is in part due to inhibitory influences from myelin and chondroitin sulfate proteoglycans (CSPGs). Peripheral nervous system (PNS) neurons exhibit increased regenerative ability compared to CNS neurons, even in the presence of inhibitory proteins. Previously, we identified over a thousand genes preferentially expressed in PNS neurons relative to their CNS counterparts. Some of these genes may account for the intrinsic differences that give the PNS enhanced regenerative ability. Primary CNS (cerebellar) neurons were transfected with plasmids encoding each of these PNS genes, to assess their ability to enhance neurite growth on a CSPG substrate. Using High Content Analysis, we evaluated the phenotypic profile of each neuron to extract meaningful data for over 1000 genes. Known growth associated proteins, such as BDNF, FGFR, PDGFR, and RelA significantly potentiated neurite growth on laminin. Interestingly, several novel genes were identified that promoted neurite growth on CSPGs (GPX3, EIF2B5, RBMX). Bioinformatics approaches also uncovered novel gene families that significantly altered neurite growth of CNS neurons.

Supported by NIH (HD057632, NS059866), U.S. Army (W81XWH-05-1-0061), Paralyzed Veterans of America, Florida JE King Biomedical Res Program and the Buoniconti Fund. VL holds the Walter G. Ross Distinguished Chair in Developmental Neuroscience.

ANALYSIS OF GENE REGULATION AND CELL FATE FROM SINGLE-CELL GENE EXPRESSION PROFILES IN C. ELEGANS

Xiao Liu[1], Fuhui Long[2], Hanchuan Peng[2], Sarah J Aerni[3], Gene Myers[2], Stuart K Kim[1]

[1]Stanford University, Developmental Biology, Stanford, CA, 94305, [2]Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA, 20147, [3]Stanford University, Computer Science, Stanford, CA, 94305

To understand gene regulatory networks in development, we developed an automatic image analysis system to measure the expression patterns of fluorescent reporters at the resolution of single cells in C. elegans. We generated expression profiles of 93 genes in 363 specific cells from L1 stage larvae, and were able to quantitatively analyze expression of each gene as well as the molecular expression signature for each cell. Expression of most genes was strongly correlated with cell fate. However, expression of some genes was correlated with cell lineage independent of cell fate. Some cells with identical fate expressed different sets of transcription factors based on their lineage. We used the molecular signatures to find repeating cell fate modules within the cell lineage and to create a molecular differentiation map, which shows points in the cell lineage when developmental fates of daughter cells begin to diverge from each other.

LARGE-SCALE LEARNING AND COMPARISON OF CELLULAR PHENOTYPES FROM IMAGES

Vebjorn Ljosa[1], Piyush B Gupta[2,3], Thouis R Jones[1], Eric S Lander[1,2,3,4], Anne E Carpenter[1]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, 02142, [2]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, [3]MIT, Dept. of Biology, Cambridge, MA, 02139, [4]Harvard Medical School, Dept. of Systems Biology, Boston, MA, 02115

Microscopy-based high-throughput screens can provide a broad view of biological responses and states at the resolution of single cells. Some phenotypes are readily identifiable; for instance, mitotic arrest can be detected by measuring the intensity of a fluorescent marker for mitosis. Other phenotypes, while apparent upon visual inspection, are much harder to identify computationally. As an example, when signaling pathways related to cell migration are stimulated, T47D breast-cancer cells take on a motile appearance, but this phenotype is not easily captured in a sparse set of measurements. Classifiers trained on hand-curated training sets can identify such phenotypes.

We present a method that can learn to recognize phenotypes without requiring hand-labeled cells for training. Instead, a classifier is learned from larger portions of the experiment known to be enriched (if only slightly) by the phenotype of interest. As an example, we use an RNAi screen of T47D breast-cancer cells. The screen was performed in duplicate, and the second replicate was treated with a protein stimulant of cellular migration. As a result, a migratory phenotype putatively related to metastasis was slightly more prevalent in the stimulated replicate (~55% vs. ~45%). Such noisy training sets are unsuitable for most machine-learning methods, but large-scale machine learning based on random features allows us to overcome the noise by using huge training sets (in our case, the millions of cells found in each replicate). As a result, a classifier specific for the response of cells to the stimulant can be created without manual classification of cells. The large-scale machine learning is made possible by projecting the cellular measurements into a random feature space constructed in such a way that the inner product in the random feature space approximates the Gaussian-RBF kernel in the original feature space.

Next, we show that inner products of population means in the random-feature space approximates the expected-likelihood kernel in the original feature space. This property makes it possible to compare heterogeneous populations of cells efficiently, approximating a comparison of the populations' kernel-density estimates in the original feature space. We demonstrate the method by clustering a collection of drugs based on their effect on cancer cells.

A HIERARCHICAL STATISTICAL MODEL FOR C. ELEGANS LINEAGE TRACING

Daniel L Mace, Thomas Boyle, Robert H Waterston

University of Washington, Genome Sciences, Seattle, WA, 98195

Expression data of developing organisms from high throughput/high resolution microscopy images presents many difficulties for computational analysis. While generic off-the-shelf methods have been adapted to extract information from the images, they are often prone to errors as they lack much of the constraints of the developing organism. We present a method for cell lineage tracing in *C. elegans* that overcomes these limitations by incorporating domain specific information into the analysis in the form of a hierarchical statistical model. Our model combines prior information about the development of the organism (e.g., cell number, division times, cell movements) with a more formal probabilistic cell nucleus detector; through the use of reversible jump markov chain monte carlo methods, we can obtain the most probable trace of the *C. elegans* lineage. With minor post-processing corrections, this model allows us to track and quantify expression through the 550 cell stage of development, much further than the previous 350 cell limitation.

IMAGE ACQUISITION AND ANALYSIS FOR HIGH THROUGHPUT PLANT ROOT MICROSCOPY

Bradley T Martsberger[1], Bradley Moore[1], Dan Mace[2], Wolfgang Busch[3], Philip N Benfey[1,3], Uwe Ohler[1,2,4,5]

[1]Duke , Institute for Genome Science and Police, Center for Systems Biology, Durham, NC, 27708, [2]Duke, Computational Biology and Bioinformatics, Durham, NC, 27708, [3]Duke, Biology, Durham, NC, 27708, [4]Duke, Computer Science, Durham, NC, 27708, [5]Duke, Biostatistics and Bioinformatics, Durham, NC, 27708

Novel high-throughput high-resolution microscopy experiments require high performance image processing for data acquisition and analysis. We have developed an automated software-based platform for plant root imaging called the RootArray. The RootArray is a unique well plate designed for the *in vivo* confocal fluorescence microscopy of up to sixty-four *Arabidopsisthaliana* roots growing in a shared environment. Root morphology is marked using a red fluorescent dye taken up in the cell walls, and each individual root expresses a green fluorescent protein (GFP) reporter for a gene of interest. The well plate provides a distinct challenge for fast image acquisition- the roots have the freedom to grow throughout the relatively large volume of the growth chamber, and continue to grow between time points. To solve this problem, we have developed a novel trainable multi-scale factor graph segmentation algorithm to detect the positions of the roots at low resolution for subsequent high resolution imaging. For the analysis of gene expression patterns (of the reporter gene), we map pixel-based GFP intensity values to individual tissue types in each root. This requires segmentation of large (100-400 megavoxels is typical) quasi-3D images. We first exploit the radial symmetry of the root to extract a representative 2D surface containing all tissue types (the medial section). This surface is then registered to a 2D atlas image. In the past year, with a single microscope, we have automatically acquired and registered over 1000 distinct roots, including over 4000 varying time points.

MULTIMODAL IMAGING AND AUTOMATED CELL-CYCLE PHENOTYPING OF DIVIDING *SACCHAROMYCES CEREVISIAE* CELLS

Michael B Mayhew[1], Nathan C Sheffield[1], Sarah Jung[2], Steven B Haase[2,4], Alexander J Hartemink[3,4]

[1]Duke University, Computational Biology & Bioinformatics Program, Durham, NC, 27708, [2]Duke University, Department of Biology, Durham, NC, 27708, [3]Duke University, Department of Computer Science, Durham, NC, 27708, [4]Duke University, Center for Systems Biology, Durham, NC, 27708

The eukaryotic cell cycle involves morphological changes over space and time as the cell replicates its genome, partitions the replicated contents evenly between parent and offspring cells, and divides. In the case of the budding yeast, *Saccharomyces cerevisiae*, these changes include myosin ring formation between the dividing cell and the nascent offspring cell or bud, emergence of the bud, mitotic spindle formation following spindle pole body (SPB) duplication and separation, nuclear migration to the myosin ring, nuclear division, and dissolution of the myosin ring, marking the end of cytokinesis.

With the aid of multiple imaging modalities, namely differential interference contrast (DIC) and fluorescence microscopy, morphological cell-cycle events can be monitored and analyzed. Analysis of images of each modality in isolation as well as modality-inherent properties complicate image processing and reduce cellular phenotyping accuracy (Kagalwala & Kanade, 2003; Li & Kanade, 2009). To overcome these issues, one could process an image of one modality by borrowing information from other modalities. For example, DAPI-stained nuclei from a blue fluorescent image could be used to pinpoint regions for DIC segmentation. Incorporation of more domain-specific knowledge (eg. expected size, shape, and spatial configuration of segmented regions or features) would also improve image processing and cellular phenotyping performance.

Here we present a DIC and fluorescence microscopy imaging platform as well as a computational image analysis pipeline, leveraging cross-modality image information and prior knowledge of the biological system. We apply the platform and pipeline to samples of dividing *S. cerevisiae* cells collected over time. We created fusion proteins to track cellular myosin ring (Myo1-mCherry) and SPB (Spc42-EGFP) status and used fluorescent stains (DAPI) to distinguish nuclei. We track bud formation and growth by DIC imaging. Each time point sample yields multiple fields of view, and each field of view is imaged multimodally using DIC as well as blue, red, and green spectral channels. The pipeline enables automatic cell segmentation, registration of DIC and fluorescent images, and identification of morphological features of cell-cycle progression. We have implemented the pipeline in MATLAB (www.mathworks.com).

THE GEOMETRY OF PHENOTYPE SPACES

Gregoire Pau, Wolfgang Huber

European Molecular Biology Laboratory, Genome Biology, Heidelberg, 69117, Germany

Many biological questions arising from high-throughput imaging screens concern the identification of novel and similar phenotypes. To achieve this task, cell phenotypes are usually quantified into high-dimensional phenotypic profiles, defining a phenotype space. Standard geometric tools, such as linear transformation and projection, are then used to determine subspaces that are relevant to address a biological question. However, the phenotype space is often anisotropic and unsupervised linear methods, such as Mahalanobis distance and principal component analysis, are unsatisfying and may report irrelevant phenotypic similarities.

To address these issues, we present two approaches relying on supervised learning methods. We first introduce distance metric learning to build a phenotypic similarity measure that is enriched for functional relationships. In the context of a genome-wide RNAi screen in HeLa cells, the method allowed us to associate uncharacterised genes to biological processes. We then propose the Lasso regression to determine the axes in the phenotype space that are relevant to discriminate complex phenotypes, based on a manually annotated training set. The method is used to investigate the differential spatial distribution of proteins in yeast upon gene deletion.

GENETIC INTERACTION ANALYSIS BY RNAI IDENTIFIES *DROSOPHILA CKA* AS A NOVEL REGULATOR OF RAS/MAPK SIGNALING

Thomas Sandmann*[1], Thomas Horn*[1], Bernd Fischer*[2], Elin Axelson[2], Wolfgang Huber[2], Michael Boutros[1]

[1]German Cancer Research Center, Division Signaling and Functional Genomics, Heidelberg, 69120, Germany, [2]EMBL, Genome Biology Unit, Heidelberg, 69117, Germany

Studies in yeast and bacteria have taken advantage of collections of deletion strains to construct matrices of quantitative interaction profiles and infer gene function. Here, we report a robust method to identify genetic interactions in metazoan tissue culture through RNAi. By performing more than 70,000 pairwise perturbations of signaling factors, we identified >600 interactions affecting different quantitative phenotypes of *Drosophila melanogaster* cells. Computational analysis of this interaction matrix allowed us to reconstruct signaling pathways and identify *Cka* as a novel, conserved regulator of Ras/MAPK signaling *in vitro* and *in vivo*.

*authors contributed equally

A HYBRID BLOB-SLICE MODEL FOR ACCURATE AND EFFICIENT DETECTION OF FLUORESCENCE LABELED NUCLEI IN 3D

Anthony Santella, Zhirong Bao

Sloan Kettering Insititute, Developmental Biology, NY, NY, 10065

Optically sectioned images of fluorescent labeled nuclei are ubiquitous in experiments studying the behavior of cells. Accurate detection of nuclei is critical for any attempt to automate quantitation in such images; nuclear position is a precursor to continuous cell tracking, and provides seed locations for segmentation of cell boundaries. Variability in nuclear appearance and under sampled volumetric data make a general solution to detection challenging. Relatively simple methods such as thresholding, or matched filters have relatively high error rates on realistically challenging images. On top of this, efficiency is critical when processing potentially terabytes of image data with tens of thousands of cells.

We present a general modular framework for the task of nuclear detection. In our approach nuclear boundary extraction is used to refine the detection results of a matched filter, leveraging simple methods to efficiently produce accurate results. Our method subdivides the problem of nuclear extraction, first segmenting nuclear slices within each 2D image plane, then using a shape model to assemble these slices into 3D nuclei around seed locations. This hybrid 2D/3D approach allows accurate accounting for nuclear shape but exploits the clear 2D nuclear boundaries that are present in sectional slices to avoid the computational burden of fitting a complex shape model to volume data. This nuclear segmentation can be used to identify both overlooked and redundantly detected nuclei. When tested over C. elegans, Drosophila, zebrafish and mouse data, our method yielded 0 to 3.7% error, up to six times more accurate and 30 times faster than published performances.

Because our approach is specialized for the characteristics of optically sectioned nuclear images, it can achieve superior accuracy in significantly less time than other approaches. Both of these characteristics are necessary for practical analysis of overwhelmingly large data sets where processing must be scalable and where the time cost of manual error correction makes it impossible to use data with high error rates. Our approach is fast, accurate, available as open source software and its learned shape model is easy to retrain. These characteristics will enable novel experimental methods utilizing complex data sets.

SEMI-SUPERVISED LEARNING FOR JOINT ANALYSIS OF TEMPORAL AND SPATIAL GENE EXPRESSION

Alexander Schliep[1], Ruben Schilling[2], Ivan G Costa[3]

[1]Rutgers University, Computer Science and BioMaPS Institute, Piscataway, NJ, 08854, [2]Max Planck Institute for Molecular Genetics, Computational Molecular Biology, Berlin, 14195, Germany, [3]Federal University of Pernambuco, Center of Informatics, Recife, Pernambuco, Brazil

Time-courses of gene expression are routinely used to detect functional modules of co-expressed genes or regulatory networks. Given the number of genes, the noise level of expression measurements, the typically low number of time-points and, often, a lack of replicates, the inferred interactions frequently arise due to chance and are not necessarily indicative of true interactions. In situ hybridization experiments with appropriate image analysis can provide the missing spatial information. However, imaging data is more costly then microarrays and not very suitable for deriving quantitative measurements of gene expression, rather giving a qualitative measurement of the spatial expression territories. Indeed, frequently image databases are manually annotated with textual descriptions of the anatomical regions exhibiting expression using ontologies of anatomical terms and approximate quantifiers.

Combining both sources of data opens the possibility to leverage the expensive imaging data by imposing spatial consistence of either clustering or network inference. Semi-supervised learning provides an appropriate machine learning framework which can help to make the most of two data sources of imbalanced abundance. Indeed, even few images can drastically improve the quality of solutions obtained for example from clustering with respect to their spatial consistence.

We present results of combining gene expression measurements during the development of the Drosophila melanogaster and in situ RNA hybridization images for different stages of its embryo. Using a semi-supervised approach we can find clusters of genes exhibiting spatio-temporal similarities in expression, or syn-expression. The temporal gene expression measurements are taken as primary data for which pairwise constraints are computed in an automated fashion from raw in situ images without the need for manual annotation. We investigate the influence of these pairwise constraints in the clustering and discuss the biological relevance of our results. Spatial information contributes to a detailed, biological meaningful analysis of temporal gene expression data. Semi-supervised learning provides a flexible, robust and efficient framework for integrating data sources of differing quality and abundance. We also present preliminary results extending this approach to further model organisms, time-inhomogeneous constraints and further inference tasks.

HIGH-THROUGHPUT ASSAYS FOR YEAST NATURAL VARIANTS

A Scott[1], P Ruusuvuori[1,2], C Ludlow[1], T Gilbert[1], G Cromie[1], Z Tan[1], V Ahyong[3], M Oeser[4], N Sakhanenko[1], N Flann[5], D Galas[1], I Shmulevich[1], A Dudley[1]

[1]Institute for Systems Biology, Seattle, WA, 98103, [2]Tampere University of Technology, Signal Processing, Tamper, FI-33720, Finland, [3]UC San Francisco, Biotechnology, San Francisco, CA, 94143, [4]University of Washington, Pharmacology, Seattle, WA, 98195, [5]Utah State University, Computer Science, Logan, UT, 84322

Understanding how differences in DNA sequence lead to phenotypic variation is one of the fundamental challenges of modern biology. This genotype-phenotype relationship is further complicated by the fact that traits are often conferred by multiple genes that interact with each other and with the environment through complicated interdependencies. Systems genetics employs large-scale, high throughput methods to decipher the network of gene function and genetic interactions.

Work in our lab takes advantage of natural variation as a rich source of genetic perturbation. We are developing automated methods for generating large numbers of recombinant progeny from crosses between natural isolates of *Saccharomyces cerevisiae*. Our goal is to marry the accuracy and detail of a conventional genetic screen with the scale of a high throughput study. A key component of this system is the ability to rapidly score thousands of progeny strains for dozens of phenotypes. By robotically pinning strains onto trays of solid media and using a simple imaging system, we can classify growth, color, and colony morphology. We measure growth as a function of the area that a colony occupies over time. A linear regression based filtering method identifies contaminants and removes them from analysis. Numerical representations of color (i.e. hue and lightness) express color as continuous variables that can be used as quantitative traits.

One characteristic that we have examined in detail is an intricate colony morphology trait known as the “fluffy” phenotype[1]. This trait is generally absent from laboratory strains, but has been observed in a variety of natural yeast isolates. Interestingly, different recombinant progeny (“siblings”) have different shapes, but genetically identical clones have similar shapes, a result that suggests that specific morphologies are genetically determined. We are using high resolution time-lapse imaging and automated image classification to study this process, which is relevant to biofilm formation and cellular differentiation. We are also using this data to develop a multi-scale model of how the behavior of single cells can lead to the emergence of large-scale structures.

[1] Palková, Z. and Váchová, L. (2006), Life within a community: benefit to yeast long-term survival. FEMS Microbiology Reviews, 30: 806–824.

QUANTITATIVE CO-LOCALISATION FOR HIGH CONTENT SCREENING APPLICATIONS

Vasanth Singan[1,2], Thouis Jones[3,4], Kathleen Curran[2], Jeremy C Simpson[1]

[1]University College Dublin, School of Biology and Environmental Science, Dublin, 4, Ireland, [2]University College Dublin, School of Medicine and Medical Science, Dublin, Ireland, [3]Broad Institute, Imaging Platform, Cambridge, MA, 02142, [4]Institut Curie, Departement du Transfert, Paris, 75248, France

Automated high content screening (HCS) platforms are capable of rapidly generating thousands of images. One of the main challenges in HCS image analysis is to ensure that these data are only subjected to minimal manual intervention such that quantification remains accurate. In the cellular context the relative distribution of proteins across different subcellular structures and organelles is a key parameter for further understanding cell function. In this regard, co-localisation analysis is potentially a very powerful tool that can be employed in the HCS arena. Existing algorithms for quantitative co-localisation either look at intensity correlation or co-occurrence. Here we present a novel algorithm that integrates intensity correlation and co-occurrence, thereby greatly improving the quality of quantification. The algorithm uses a non-parametric rank-based weighting of channels to integrate correlation and co-occurrence. We have successfully implemented this algorithm in a commercial HCS analysis software (Olympus Scan^R Analysis), in addition to open source software (CellProfiler). We present results of a fully automated high-throughput experiment acquired using Scan^R Acquistion and processed using Scan^R Analysis and CellProfiler.

ILASTIK: INTERACTIVE LEARNING AND SEGMENTATION TOOL KIT

Christoph Sommer, Christoph Straehle, Ullrich Koethe, Fred A Hamprecht

University of Heidelberg , HCI, Heidelberg, 69115, Germany

The analysis of biological high throughput images often requires segmentation as a first step. "Segmentation" is the process of partitioning digital images into meaningful regions. We propose *ilastik* as an easy-to-use tool which allows the user to perform segmentation and identification in a unified way without expertise in image processing.

The tool learns from labels provided by the user through a convenient mouse interface. Based on these labels, ilastik infers a problem specific segmentation. A *random forest* classifier is used in the learning step, in which each pixel's neighborhood is characterized by a set of generic (nonlinear) features. Ilastik supports up to 3D multi-spectral data and makes use of all 4 dimensions in the feature calculation. The user can choose among preselected feature groups (color, texture, etc). A plug-in system offers extensibility for more specific questions.

Ilastik provides a real time feedback of the classifier that enables the user to interactively refine the segmentation result. An uncertainty measure guides the user to difficult regions in the images. In order to achieve real time performance, the underlying algorithms are parallelized to exploit the capabilities of modern multi-core machines. Once a classifier has been trained on a set of representative images, it can be exported and used to automatically process a very large number of images (e.g. using the CellProfiler pipeline, to which it is compatible).

The proposed method is still limited by heavy spatial overlap of objects, and cannot solve tasks that require a higher-level semantic understanding of the images; however, for many standard tasks it performs surprisingly well.

We will illustrate the performance of ilastik in a live demonstration, and show quantitative results on 2D and 3D real world multichannel data. The demonstration will show that ilastik is able to solve standard tasks that previously would have required manual tuning or even hand-tailored algorithms.

STOCHASTIC FATE CHOICE IN A MULTI-CELLULAR ORGANISM: PATTERNING THE DROSOPHILA MELANOGASTER RETINA

Pranidhi Sood[1], Robert Johnston[2], Claude Desplan[2], Edo Kussell[1]

[1]New York University, Center for Genomics and Systems Biology, New York, NY, 10003, [2]New York University, Developmental Genetics, New York, NY, 10003

There are two general models that describe the acquisition of cell fate: a deterministic one in which a cell's identity is determined by its environment and lineage and a stochastic one in which a cell's fate is determined by noise, for example in gene expression. Studies of the latter type have largely been performed in unicellular organisms, for example the study of ComK in B. subtilis. Here, we study stochastic fate acquisition in a multicellular organism, specifically in the development of the photoreceptor mosaic of the retina of D. melanogaster. Using 3D confocal images of the retina in both wild type and mutant backgrounds, we will characterize the distribution of light-sensing rhodopsins which distinguish photoreceptor types. This study is unique in that we will quantitatively characterize a stochastic fate in a multicellular organism that plays a physiological role in the perception of color.

TRAINING BOUNDARY DETECTORS FOR SEGMENTATION BY LEARNING MINIMAX DISTANCES

Srinivas C Turaga[1], Kevin L Briggman[2], Moritz N Helmstaedter[2], Winfried Denk[2], H Sebastian Seung[3]

[1]MIT, Brain & Cognitive Sciences, Cambridge, MA, 02139, [2]Max-Planck Institute for Medical Research, Biomedical Optics, Heidelberg, D-69120, Germany, [3]MIT/HHMI, Brain & Cognitive Sciences, Cambridge, MA, 02139

Many algorithms for image segmentation rely on boundary detection as a first step. The output of a boundary detector (BD) is then used to group voxels into regions corresponding to objects. While hand-designed detectors based on image derivatives and other local features have found great utility, in recent years machine learning (ML) has been used to solve this problem. A segmentation generated by a human expert is used to train a classifier to predict boundaries based on image patches. ML-based BDs have the advantage of being able to automatically learn the image features that are diagnostic of boundary voxels.

ML algorithms can contain many tens of thousands of free parameters which are adjusted by a training procedure to minimize an error measure. BDs are usually trained to minimize the number of voxels incorrectly classified as boundary or not. With powerful ML architectures such as convolutional networks (CN), excellent BDs can be trained in the sense that the number of misclassified voxels is very low (Jain et al. 2007; Turaga et al. 2009). However this error measure is misleading since misclassification of just a few voxels can lead to dramatically different segmentations. A narrow process can be split in two if a few voxels along it are mistakenly classified as boundary. Two neurites can be merged into one, if there is even a single voxel-wide gap in the boundary between them.

We present a new error measure that is sensitive to split and merge mistakes and show how this error measure can be trained to produce BDs leading to good segmentations. Our learning algorithm casts image segmentation as the problem of classifying the connectivity of pairs of image pixels. Rather than asking whether a given voxel must be classified as boundary according to the human segmentation, we ask if pairs of voxels should be connected or not by classifying a given voxel as boundary or not. This enables us to train a BD that is superior in the sense of generating segmentations with low split and merge rates, rather than in the less useful sense of the number of correctly classified boundary voxels. We show how this training algorithm relates to the learning of a particular kind of shortest path known as the minimax shortest path between pairs of voxels.

We demonstrate the superiority of voxel-pair-wise minimax learning over conventional voxel-wise boundary training for the task of segmenting neurons in 3d SBFSEM (Denk & Horstmann 2004) imagery using a CN.

IMAGE-BASED HIGH-THROUGHPUT SCREENING USING *C. ELEGANS*

Carolina Wählby[1], Zihan Hans Liu[1], Tammy Riklin-Raviv[2], Lee Kamentsky[1], Katherine Madden[1], Vebjorn Ljosa[1], Annie L Conery[3], Eyleen O'Rourke[3], Javier E Irazoqui[4], Polina Golland[2], Frederick M Ausubel[3], Anne E Carpenter[1]

[1]Broad Institute, Imaging Platform, Cambridge, MA, 02142, [2]MIT, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, 02139, [3]Harvard Medical School, Dept. Molecular Biology, Boston, MA, 02144, [4]Harvard Medical School, Dept. of Pediatrics, Boston, MA, 02114

As automated microscopy and image analysis has improved, screens based on cultured cells have proven to be powerful and versatile. However, biological pathways that involve organ systems and multi-cellular interactions, such as metabolism and infection, can be better studied in whole animals. The roundworm Caenorhabditis elegans is a well studied and effective model system for many biological processes, and recent advances in robotic sample preparation have made it possible to use *C. elegans* in high-throughput screens.

Screening in high throughput limits image resolution and time-lapse information. Still, the images are rich in information and the number of images for a standard screen often exceeds 100 000, ruling out visual inspection. For many screens, it is possible to find 'hits' by looking at image averages, such as total fluorescent stain per image area covered by worms. However, worm population heterogeneity calls for image analysis approaches that extract information on a per-worm basis. Each image (i.e., well) of a screen typically holds 15 worms, and apart from robustness to noise and variations in illumination, the largest challenge from an image analysis point of view is clustering of worms, which prevents extraction of descriptive features from individual animals. We present a novel approach for the extraction of individual *C. elegans* from clusters. Our approach is based on constructing a low-dimensional shape-descriptor space and applying a graph-based search to untangle clusters. Once individual worms are extracted, simple features such as shape, intensity and texture can be extracted to discriminate between live and dead worms or quantify fat accumulation. We also show that worms can be 'straightened' and aligned in order to compare staining patterns of biologically relevant markers.

NOVEL SEGMENTATION ALGORITHMS FOR DIFFERENTIAL INTERFERENCE CONTRAST MICROSCOPY IMAGES

Quanli Wang[1,2]

[1]Duke University, Duke Center for Systems Biology, Durham, NC, 27708,
[2]Duke University, Department of Statistical Science, Durham, NC, 27708

Differential Interference Contrast (DIC) microscopy images are notoriously difficult to automatically segment in spite of their wide appeal due to the pseudo-3D appearance that can be easily interpreted by human observers. Conventional algorithms for image processing often use threshold-based criteria to identify objects of interest; such methods usually fail for DIC images as they have a range from bright-to-dark with intermediate intensity values that are very close to the background. Recent development of model based algorithms show promise for DIC image reconstruction, but are inherently slow and typically require substantial customization and parameter tuning. With a view to effective segmentation and data extraction in routine application across multiple contexts in single cell studies, we have developed a direct approach to DIC image segmentation that shows substantial benefits in compute, parameter insensitivity and resulting accuracy. We will demonstrate this algorithm use both static DIC images and a time-lapse movie
from single-cell studies.

RAPID AND ACCURATE PHENOTYPING OF EMBRYONIC LETHALITY VIA IMAGE ANALYSIS OF *C. ELEGANS* DEVELOPMENTAL STAGES FROM HIGH-THROUGHPUT IMAGE DATA

Amelia G White[1,2], Huey-Ling Kao[1], Patricia G Cipriani[1], Eliana Munarriz[1], Davi Geiger[3], Kris C Gunsalus[1], Fabio Piano[1]

[1]New York University, Department of Biology, New York, NY, 10003, [2]Rutgers University, Department of Computational Biology and Molecular Biophysics, Piscataway, NJ, 08854, [3]New York University, Department of Computer Science, New York, NY, 10012

We present an image analysis system (DevStaR) for automatic classification of the developmental stages of *C. elegans* animals from a population of mixed stages. DevStaR enables efficient high throughput quantification of *C. elegans* embryonic lethality and has the potential to be extended to other phenotypes. DevStaR is an object recognition machine comprising several hierarchical layers that build successively more sophisticated representations of the objects to be classified. The algorithm segments the objects, decomposes the objects into parts, extracts features from these parts, and classifies them using an SVM and global shape information. This enables correct classifications in the presence of complicated occlusions and deformations of the animals.

We are currently using this system to analyze phenotypic data from *C. elegans* high-throughput genetic screens, and have used it to process over 500,000 images for lab users so far. To assess the performance of DevStaR, we have used it to measure embryonic lethality in populations of an embryonic lethal temperature sensitive strain of *C. elegans* and compared the results to manual counts of brood size at four temperatures ranging from the restrictive to the permissive temperature. We find that the embryonic lethality measured by DevStaR is very similar to results from manual counting at all temperatures.

Our system overcomes a previous bottleneck in image analysis by achieving near real-time scoring of image data in a fully automated manner. Our system reduces the need for human evaluation of images and provides rapid quantitative output that is not feasible at high throughput by manual scoring. This system can be readily adapted for *C. elegans* screening scenarios in other laboratories. Moreover, the methods used could be usefully applied to solve other high-throughput imaging problems.

3D VOLUMETRIC EX-VIVO MOUSE EMBRYO IMAGING AND IMAGE REGISTRATION USING MRI, MICROCT AND OPTICAL PROJECTION TOMOGRAPHY

Michael D Wong, R M Henkelman

Mouse Imaging Centre, Hospital for Sick Children, Toronto, M5T 3H7, Canada

Introduction

There is a world wide effort to knock out each of the 23,000 genes of the mouse, one at a time, and breed the resultant mice. Thirty percent of these mouse lines are embryonic lethal, accentuating the need for mouse embryo phenotyping. 3D volumetric high resolution imaging is one method to characterize morphologic phenotypes in both wild-type and mutant mouse embryos. This study outlines the imaging software, hardware, and protocols for ex-vivo mouse embryo imaging using Optical Projection Tomography (OPT), Magnetic Resonance Imaging (MRI), and X-Ray MicroCT (μCT) along with computer automated image registration of the resultant images.

Imaging

OPT employs ultraviolet light to excite endogenous fluorophores in a small semi-transparent sample like the developing mouse embryo and the resultant emitted photons are counted by a charged coupled device (CCD) camera. 3D volumetric images are acquired in half an hour with an image voxel size as small as 1.3 microns isotropic.

μCT typically produces inadequate soft tissue contrast in the mouse embryo, but immersing the mouse embryo in 0.1N potassium iodine for 5 hours generates remarkable soft tissue contrast in a 15 day old (E15.5) mouse embryo. An image of seven E15.5 mouse embryos can be acquired in three hours with a voxel size of 30μm or a single embryo data set can be generated in three hours with an isotropic voxel size of 14μm.

MRI exhibits superior soft tissue contrast in the mature mouse embryo (E15.5) using gadolinium as a contrast agent. An array of 3 solenoid transmit-receive MRI coils is used to image three E15.5 mouse embryos in parallel over the course of an overnight scan with an isotropic resolution of 25μm without zero-filling or filtering.

Methods

Six E15.5 mouse embryos were imaged with both μCT and MRI while six E12.5 mouse embryos were imaged with all three imaging modalities. The 6 mouse embryo images underwent computer automated image registration and the image quality of the population average was used to qualitatively validate the usefulness of the three modalities for computer automated analysis and phenotyping.

Results

All three imaging modalities generate 3D volumetric mouse embryo images with adequate resolution, signal-to-noise, and contrast that our computer automated image registration algorithm can be used to produce high-quality population averages of a series of mouse embryo images.

Discussion

OPT, MRI and μCT imaging parameters and mouse embryo sample preparation for each of these modalities has been optimized for high-throughput imaging with a focus on compatibility with automated image registration.

USE OF ADHESIVE MICROPATTERNS IN HCS ACCELERATES IMAGE ACQUISITION AND ANALYSIS, INCREASES SENSITIVITY OF DETECTION AND PROVIDES STATISTICALLY SIGNIFICANT DATA WITH FEWER CELLS.

Joanne Young[1], Sébastien Degot[1], Muriel Auzan[1], Violaine Chapuis[1], Anne Béghin[2], Alexandra Fuchs[1]

[1]CYTOO, Cell Architects, Grenoble, 38040, France, [2]CCQ, Faculté de Médecine, Lyon, 69008, France

HCS is often carried out with adherent cells grown on homogeneous surfaces. Cells can divide, move and spread in all directions leading to variability in cell shape, behavior and a heterogeneous response of cells to stimuli. The variation in cell geometry also slows down automatic image acquisition and analysis (eg. image segmentation) and complicates quantification of changes in targets such as the cytoskeleton. These bottlenecks in HCS can be resolved by replacing uniform coatings with micropatterned surfaces.

Using L-shaped micropatterns that highlight actin bundles present at the cell's free edge, we developed a cell-based assay that quantifies drug impact on the cytoskeleton and is compatible for high throughput phenotyping. We present data on 4 drugs at 5 different concentrations in 96-well plate format. The drugs target actin directly (blebbistatin, Y27632, cytochalasin D) or indirectly (nocodazole) causing a decrease or increase in cell contractility respectively. Both phenotypic changes were quantified using an automated image analysis macro, HypothenuseEDM™ that relies on an Euclidean Distance Map function for categorizing cells as intact or collapsed.

Because micropatterns display each cell in a reproducible fashion with normalized internal organization, a marked increase in assay sensitivity was obtained. For example, the effect of Y27632 on cell tension was detected at a dose one hundred times less than commonly used in cellular assays. Furthermore, statistically significant results ($p<0.005$) were obtained by analyzing as few as 50 cells, favoring use of micropatterns for miniaturization to 1536-well plates and screening of precious primary or stem cells. A useful visualization tool, the Reference Cell™ which is a heat-map representation of "the average cell" response to perturbation further increases efficiency allowing single measurements on a population of cells, providing a standard for comparing between experiments and enables display/quantification of mechanistic differences in drug modes of action.

Micropatterns thus offer a clear advantage over conventional culture wells for HCS simplifying automated imaging, processing and analysis, increasing sensitivity of readouts and minimizing the number of cells for analysis. Finally, careful design of micropattern shape creates possibilities for screening new cellular parameters, as shown by our novel cell-based assay that measures alterations in cytoskeleton tension in an automated setting.

A NOVEL PHENOTYPIC DISTANCE MEASURE FOR IMAGE-BASED HIGH-THROUGHPUT SCREENING

Xian Zhang[1], Gregoire Pau[2], Wolfgang Huber[2], Michael Boutros[1]

[1]German Cancer Research Center (DKFZ) and University of Heidelberg, Division of Signaling and Functional Genomics, Heidelberg, 69121, Germany, [2]EMBL, Genome Biology Unit, Heidelberg, 69117, Germany

Image-based high-throughput screening investigates complex phenotypes under individual RNAi or small molecular treatments, and facilitates gene network dissection and drug candidate identification. Microscopy readouts generated from these screens are typically quantified via extracting image-based features for individual cells, which results in large-scale multidimensional data sets. How to interpret the data sets to reveal biologically interesting information remains challenging. Dimension reduction and statistical learning methods have been previously proposed, but they often consist multiple steps and parameters, which prevents easy adaptation to other screening experiments. Additionally, to our knowledge these methods have not been evaluated on a common screening data set.

Here we propose a new analysis method, which computes the phenotypic distance between treatments via cell classification and cross validation. One can identify phenotypes based on phenotypic distances between samples and controls, and cluster phenotypes based on phenotypic distances between samples. The proposed method, together with previous methods, is evaluated on a common screening data set, with quality metrics addressing replicate reproducibility, control separation and clustering enrichment.

The proposed method shows a better overall performance compared with previous methods in the tested screening data set, and requires fewer parameters which facilitates adaptation to other screening data sets. Phenotypic distance methods and quality metrics used in this study are implemented in a freely available R/Bioconductor package, which can serve as a toolbox for data analysis of image-based high-throughput screens.

HIGH THROUGHPUT SINGLE-CELL PHARMACOLOGY USING SEMANTIC DATACUBES.

Peter K Sorger , Mario Niepel, Bjorn Millard, Jeremy Muhlich

Harvard Medical School, Dept. of Systems Biology, Boston, MA, 02115

The success of genomics in analyzing and interpreting sequence data has been possible in large part because the data are machine-readable and web-accessible. However, most non-genomic data in cell and molecular biology reside in loosely organized files of poorly documented provenance. This situation prevails, despite many attempts to rectify it, because the relational databases that are ubiquitous in genomics are ill-suited to rapidly evolving experimental designs, new data formats and changeable algorithms encountered in cell-based studies that use imaging, multiplex biochemistry, flow cytometry etc. In this paper we present a new approach to management of rich experimental data using semantically typed data hypercubes (SDCubes), and describe new open-source software (ImageRail) that uses SDCubes for high-throughput imaging. In our approach, experimental design and its day-to-day evolution determine how data are organized and stored, not rigid standards or predetermined schema. We demonstrate the practicality of our approach by applying it to the collection and analysis of multi-factorial drug dose-responses data at the single-cell level. We observe variability in the responses of tumor cells to anticancer drugs to be maximal when drug-ligand concentrations approximate physiological levels, a phenomenon potentially relevant to variability in drug response in patients.

SURVEYING THE YEAST PROTEOME USING HIGH-CONTENT SCREENING.

Yolanda T Chong, Judice Koh, Mike J Cox, Charlie Boone, Brenda J Andrews, Jason Moffat

University of Toronto, Donnelly Centre, Toronto, M5S 3E1, Canada

Many proteins are regulated through changes in both their localization and their abundance. For example, in response to nutrient availability, substrates of the TORC1 complex, Gln3, Gat1 and Sfp1, translocate between the cytoplasm and the nucleus, while ribosomal proteins are down-regulated. To comprehensively assess changes in abundance and localization in yeast proteome under chemical and genetic perturbations, we developed a high-content screening assay to survey the yeast GFP collection.
We use synthetic genetic array (SGA) technology to introduce different markers and mutations into the GFP-tagged gene collection. We then use high throughput microscopy and computational image analysis to rapidly acquire high resolution images and quantify changes in protein localization and abundance. To date, we have surveyed approximately 30 million cells and acquired 6 billion measurements which we have used to train 69 classifiers to automatically distinguish stage of the cell cycle, protein abundance and protein location. Our assessment of protein abundance in wild-type cells correlates well with published high-throughput estimates of protein abundance and our localization assignments agree with published manually annotated localization patterns, validating our classifiers and general approach. We have been able to assign patterns to proteins previously annotated as ambiguous in the GFP collection and have found over 550, greater than 2-fold, changes in protein abundance across 4 treatments and 11 genetic backgrounds. Our proof-of-principle screens demonstrate that our approach is an effective means of monitoring proteins in different genetic backgrounds and chemical conditions. The combination of large-scale genetics and high-content screening will enable the quantitative assessment of the abundance and localization of proteins in response to thousands of different genetic perturbations, providing an unprecedented means of phenotypic profiling.

IMAGING ROOTS FOR REGULATORY AND PHYSICAL NETWORK RECONSTRUCTION

Philip N Benfey[1,2], Uwe Ohler[1,2], Wolfgang Busch[1], Bradley Martsberger[1], Anjali Iyer-Pascuzzi[1,2], Chris Topp[1,2], Richard Twigg[1,2], Sophie Fang[1,2], Paul Zurek[1,2], Bradley T Moore[1], John Harer[1,4], Herbert Edelsbrunner[1,5], Yuriy Mileyko[1,4], Joshua Weitz[3], Olga Symonova[3], Ying Zheng[1,5]

[1]Duke University, IGSP Center for Systems Biology, Durham, NC, 27708, [2]Duke University, Biology Department, Durham, NC, 27708, [3]Georgia Tech, School of Biology, Atlanta, GA, 30332 , [4]Duke University, Mathematics Department, Durham, NC, 27708, [5]Duke University, Computer Science, Durham, NC, 27708

Specification and maintenance of cell identity are central processes of development. In an effort to understand the regulatory networks that control cell identity, we have profiled all cell types and developmental stages within a single organ, the Arabidopsis root. We are using computational methods to infer networks functioning within different cell types and developmental stages and have begun to test the hypothesized relationships. Our current efforts are aimed at understanding the responses to abiotic stess at high spatio-temporal resolution. We are developing new expression analysis platforms and means of analyzing 4-D data sets. We are also analyzing the dynamics of growth of physical root networks using novel non-invasive imaging methods and developing mathematical descriptors of root system architecture.

HIGH CONTENT ANALYSIS OF NEUTROPHIL PHENOTYPE DURING CHEMOTAXIS

Ivar Meyvantsson[1], Elizabeth Vu[2], Victoria Echeverria[2], Casey Lamers[1], Allyson Skoien[2], Steve Hayes[2]

[1]Bellbrook Labs, Biomedical Engineering, Madison, WI, 53711, [2]Bellbrook Labs, Cell Biology, Madison, WI, 53711

During directional migration in response to chemoattractant concentration gradients - known as chemotaxis - neutrophils undergo pronounced changes in morphology that include spreading, protrusions at the front and narrow appendages at the trailing edge. However, analysis of this phenotype has been limited in throughput by existing assay technologies, such as the Dunn chamber and under agarose assay, which can only support a handful of parallel conditions. To address this issue an assay plate with 96 microfluidic gradient generators was developed as well as fully automated methods of operation. These plates were used to study dose-dependent stimulation and inhibition of primary human neutrophil chemotaxis, and phenotypic responses to various chemo-attractant concentration profiles and inhibitors. Primary human neutrophil samples are very time sensitive and cannot be frozen for later use. Automated liquid handling, imaging and image processing enabled multiple conditions to be tested with replicates using the same biological sample. Under conditions that support chemotaxis a comparison of inhibitors that target either signaling (Wortmannin) or cytoskeletal kinetics (Latrunculin B) revealed drastically different effects on phenotype; while the signaling inhibitor reverted the cells to a non-migratory phenotype, the cytoskeletal inhibitor caused additional spreading and loss of elongation consistent with the known mechanisms of the inhibitors.

A VIRTUAL 3D ATLAS FOR QUANTITATIVE ANALYSIS OF PLANTS

Thorsten Schmidt[1,3], Taras Pasternak[2], Alexander Dovzhenko[2,3], Dorothée Aubry[2], William Teale[2], Hans Burkhardt[1,3], Olaf Ronneberger[1,3], Hartman Harz[6,7], Rainer Daum[6,7], Rainer Uhl[6,7], Klaus Palme[1,3,4,5]

[1]Albert-Ludwigs-Universität , Institute of Computer Science, Freiburg, 79110, Germany, [2]Albert-Ludwigs-Universität , Institute of Biology II, Freiburg, 79104, Germany, [3]Albert-Ludwigs-Universität , bioss – Center for Biological Sig3Albert-Ludwigs-Universität Freiburgnaling Studies, Freiburg, 79104, Germany, [4]Albert-Ludwigs-Universität , FRIAS – Freiburg Institute for Advanced Studies, Freiburg, 79104, Germany, [5]Albert-Ludwigs-Universität, FRISYS, Freiburg, 79104, Germany, [6]TILL Photonics GmbH, Gräfelfing, 79104, Germany, [7]Ludwig-Maximilians-Universität München, BioImaging Zentrum BIZ, Martinsried, 82152, Germany

Systems biology demands quantitative, comprehensive data on all aspects of biological phenomena. These data need to be mapped into the three-dimensional landscape of cell, tissues and organs. But we lack a robust three-dimensional (3D) digital atlas of positions of positions of nuclei and other cellular components in any postembryogenic stage. Such an atlas will have important applications by providing previously unavailable knowledge on cellular characteristics and allow high-throughput analysis of cellular information at single cell resolution. We describe the development of novel instrumentation for high-throughput image analysis to map the subcellular location of hundreds of RNAs and proteins in the root of Arabidopsis thaliana providing detailed knowledge of the spatio-temporal distribution of cellular constitutents critical for building three-dimensional (3D) computational models. We developed the intrinsic Root Coordinate System (iRoCS) as a reference model for the root apical meristem (RAM). An automated image processing pipeline was developed to annotate cells according to their location, type, and division status. iRoCS enabled the direct quantitative comparison between roots at single cell resolution, and was able to incorporate any recognizable feature. To demonstrate the power of the technique, we measured the capacity of changing patterns of auxin flux within the RAM to effect subtle changes in cell division patterns. We will discuss applications and innovative opportunities arising from this technological advance for crop plants.

CELLULAR HETEROGENEITY IN MODELS OF CANCER AND METABOLISM: WHICH DIFFERENCES MAKE A DIFFERENCE?

Lani Wu

UT Southwestern Medical Center, Cecil H. & Ida Green Scholar in Biomedical Computational Science, Dept. of Pharmacology & Green Ctr for Systems Biology, Dallas, TX, 75390

Phenotypic heterogeneity in cellular populations has been widely observed. However, interpreting possible roles of cellular heterogeneity in mammalian biology and disease is an exceedingly challenging problem due to the complexity of possible cellular phenotypes, the large dimension of potential perturbations, and the lack of methods for separating meaningful biological information from noise. To investigate cellular heterogeneity, our lab has developed an approach for summarizing the single-cell phenotypes of cellular populations into human- and machine-interpretable “profiles” that can be used to investigate changes due to physiological, pathological, or environmental perturbations. We will discuss recent progress in interpreting the role of heterogeneity in models of cancer and metabolism.

THE OPEN MICROSCOPY ENVIRONMENT: OPEN TOOLS FOR BIOLOGICAL IMAGE INFORMATICS

Jason Swedlow, OME Consortium

University of Dundee, WT Centre for Gene Regulation & Expression, Dundee, DD1 5EH, United Kingdom

The Open Microscopy Environment (OME) is a multi-site collaborative effort among academic laboratories and commercial entities that produces open tools to support data visualization, management, and analysis for biological imaging. Designed to interact with existing commercial software, all OME formats and software are free, and all OME source code is available under GNU public "copyleft" licenses. With a strong foundation for light microscopy, OME has begun extending its coverage to other fields of biological imaging, such as high content screening (HCS), electron microscopy, scanning probe microscopy, and digital pathology.

OME develops and releases three different components:
1. The OME Data Model (http://ome-xml.org) provides a specification for saving metadata and exchanging metadata in microscopy and HCAs. Originally, this data model treated each image as having spatial, temporal and spectral dimensions. We will report on an ‘n-dimensional data model to support more complex data structures, as seen in fluorescence lifetime microscopy.

2. The OME-TIFF file format (http://ome-xml.org/wiki/OmeTiff) and the Bio-Formats file format library (http://openmicroscopy.org/site/products/bio-formats) provide an easy-to-use set of tools for converting data from proprietary file formats into a common, open format. Bio-Formats plugs into ImageJ and Matlab, enabling access to over 90 different image file formats.

3. The OMERO platform is a Java-based server and client application suite that combines an image metadata database, a binary image data repository and high performance visualization and analysis. OMERO includes interfaces for Java, C++ and Python to support a wide variety of client applications and support for Matlab-based applications like CellProfiler. For computational analysis of microscopy or HCS images, this standardised interface provides a single mechanism for accessing image data of all types-- regardless of the original file format. OMERO contains a scripting service to support all Python-based image processing and analysis algorithms, and job distribution system to make use of clusters and other distributed computing resources. OMERO supports annotation of any data type with any kind of structured data, and any text-based structured annotations (e.g., PDF, DOC, XLS, PPT) are indexed for searching by any OMERO client. Ontologies can be represented in XML within the server and made accessible to clients via the OMERO API. OMERO is used in the Columbus® data management system (PerkinElmer, Inc.), the softWoRx® DMS system (Applied Precision, Inc.), and is the engine that runs the JCB DataViewer (http://jcb-dataviewer.rupress.org), the first publication system for original image data in the life sciences, and ASCB’s Cell Image Library (http://www.celllibrary.org).

More information is available at http://openmicroscopy.org.

IMAGE ANALYSIS FRAMEWORK FOR HIGH-THROUGHPUT PHENOTYPING

B.S. Manjunath

University of California, Santa Barbara, CA, 93106

Developing analysis workflow for high-throughput imaging pose interesting challenges and requires novel tools/frameworks for efficient processing, access (search and retrieval), and use. Analysis development also requires management of experimental metadata, ground truth as well as utilizing large amounts of image data. We present recent developments in image analysis frameworks towards building a scalable infrastructure that can leverage the emerging cloud/grid computing technologies, and describe our efforts with the iPant cyber-infrastructure in delivering integrated image analysis methods.

DATABASING CONCEPTS FOR AUTOMATED CELL IMAGING.

Anne L Plant, Talapady N Bhat, John T Elliott

National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, MD, 20899

Automated microscopy provides the opportunity to collect vast amounts of data, and increases the need to store and search large datasets. It would be of great benefit to hypothesis building and discovery if image data were easily shared between labs, could be reanalyzed to extract additional information, and could be reliably combined with data from other experimental modalities. Currently the lack of experimental details about cell imaging experiments, and the lack of quantitative benchmarks, limit the reuse and sharing of image data. To make cell image data fully useful by others, it must be accompanied by metadata terms that fully describe the relevant experimental parameters, and cell image datasets must be searchable with sophisticated queries. Ideally, complete metadata should include details of the handling and characterization of the experimental cell line prior to the experiment, treatment of experimental samples, imaging conditions, replicate information, benchmarking materials, and instrumentation details. Challenges for metadata collection include the very large number of terms that could be required and the need to be able to add additional new terms for new experiments. We suggest that the significant challenges associated with standardizing vocabulary terms may be alleviated by identifying a limited set of controlled terms, along with rules for adding new terms. We also propose a schema-less and database friendly approach to organize metadata terms that is independent of specific data formats or software tools. This approach relies on data tables of metadata, which can be semantically organized during a user-defined search query. By assembling metadata terms in a data table, image data can be searched across different datasets and different federated sites. A file-naming rule keeps image data, metadata, protocol, and other files unambiguously linked. Logical operations can be performed on the metadata, such as subtraction of metadata terms to identify differences between experimental details of different images series.

OPTICAL IMAGING AND QUANTITATIVE ANALYSIS OF BEHAVIOURAL PHENOTYPES IN *C. ELEGANS*

Eviatar Yemini, Victoria Butler, Tadas Jucikas, William Schafer

MRC Laboratory of Molecular Biology, Cambridge, MA, CB4 1JD

Understanding the relationship between genes, neurons and behaviour requires the development of methods for the rapid and consistent quantitation of behavioural phenotypes. We are developing machine vision approaches to automate the collection of video image data from *C. elegans* mutants and quantitatively analyze critical parameters related to nervous system function. We are using our vision systems to quantify the similarities of *C. elegans* mutant phenotypes and to investigate the natural clustering of mutant behavioral patterns. Ultimately, we hope to use these methods to generate a large-scale phenotypic database for *C. elegans*, which should make it possible to identify groups of mutants neuronal ablations and pharmacological treatments that have similar effects on behaviour or development, and therefore infer involvement in a common biological function.

We are also interested in understanding the neural basis for behavioural phenotypes. To address this, we have worked with Rex Kerr and Mitya Chklovskii at Janelia Farm to develop a tracking microscope suitable for in vivo calcium imaging from the neuromusculature of freely-behaving worms. Using this system, we have been able to monitor the activities of motorneurons and muscles of worms engaged in locomotion. We are using this system to investigate how alterations in behavioural patterns relate to changes in neuromuscular activity.

AUTOMATED HOME-CAGE BEHAVIORAL PHENOTYPING OF MICE

Hueihan Jhuang[1], Nicholas Edelman[1], Estibaliz Garrote[1], Tomaso Poggio[1], Andrew Steele[2], Thomas Serre[3]

[1]Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences & McGovern Institute , Cambridge, MA, 02139, [2]California Institute of Technology, Broad Fellows in Brain Circuitry Program, Division of Biology, Pasadena , CA, 91125, [3]Brown University, Department of Cognitive, Linguistic & Psychological Sciences, Brown Institute for Brain Sciences, Providence , RI, 02192

Neurobehavioral analysis of mouse phenotypes requires the monitoring of mouse behavior over long periods of time. Automating analysis could compensate the limitations of human assessment: cost, time, and reproducibility. Most previous automated systems (Noldus et al, 2001; Goulding et al, 2008) rely on the use of non-visual sensors or video tracking to monitor behavior. Such systems are suitable for studies involving spatial measurements, but not for fine behaviors such as grooming. A few vision-based systems have been described, but not been comprehensively evaluated on videos from a real lab setting (Dollar et al, 2005, Xue & Henderson, 2009) or cost prohibitive (HomeCageScan 2.0, CleverSys, Inc).

Here, we describe a *trainable, vision-based, automated high-throughput* system for the behavioral analysis of mice in their home cage. The system computes a sequence of feature descriptors for each input video based on the motion of the animal (Jhuang et al, 2007). A classifier is then trained from annotated videos to predict an output label for every frame of the video. In order to train and test our system on a real-world setting where mice behaviors are continuously observed over hours, we collected a dataset containing 10 hours of videos, where each frame is labeled as one of the 8 behaviors of interest.

Our system achieves 77.3% agreement with human labelers, which is significantly higher than the 60.9% agreement of HomeCageScan 2.0 system, and on par with the 71.6% agreement between two independent sets of human labelers. To demonstrate the applicability of the system to large-scale phenotypic analysis, we characterized the behaviors of four strains of mice: wild-derived strain CAST/EiJ, the BTBR strain (a potential model of autism, McFarlane et al, 2008), and two inbred strains, C57BL/6J and DBA/2J. The system output labels were used to validate the behavioral difference between strains and predict the strain identity with 90% of accuracy. We demonstrate the trainability of the system by training and testing the system on a set of videos containing a new set of behaviors corresponding to animals interacting with running wheels. The system achieves 93% agreement with human labelers.

The system has been published (Jhuang et al, 2010) and is now actively under development.

TOWARDS HIGH-THROUGHPUT ANALYSIS OF DROSOPHILA AGGRESSION AND COURTSHIP

Pietro Perona, Heiko Dankert, Michel Maire, Liming Wang , David Anderson
C
alifornia Institute of Technology, Pasadena, CA, 91125

We introduce a method based on machine vision for automatically measuring aggression and courtship in *Drosophila melanogaster*. The genetic and neural circuit bases of these innate social behaviors are poorly understood. High-throughput behavioral screening in this genetically tractable model organism is a potentially powerful approach, but it is currently very laborious. Our system monitors interacting pairs of flies and computes their location, orientation and wing posture. These features are used for detecting behaviors exhibited during aggression and courtship. Among these, wing threat, lunging and tussling are specific to aggression; circling, wing extension (courtship 'song') and copulation are specific to courtship; locomotion and chasing are common to both. Ethograms may be constructed automatically from these measurements, saving considerable time and effort. This technology should enable large-scale screens for genes and neural circuits controlling courtship and aggression.

USING KNOCKOUT MOUSE PHENOTYPING AND REPORTER GENE EXPRESSION ANALYSIS AS A HIGH-THROUGHPUT THERAPEUTIC TARGET VALIDATION SCREEN

Nicholas W Gale

Regeneron Pharmaceuticals, Inc, Therapeutic Target Discovery, Tarrytown, NY, 10591

The availability of the mouse genomic sequence, together with recently developed high-throughput methods for mouse genetics and mutagenesis, and large scale funding projects have rapidly accelerated the pace of generation of mutant mouse models. Systematic large-scale phenotyping efforts evaluating these models have also begun to generate information on an unprecedented scale, leading to an understanding of the roles of individual genes in biology. Along with our high throughput gene targeting efforts, funded by the NIH KOMP (Knockout Mouse Project), we have employed a novel high-volume primary phenotyping screen in which we generate large F2 cohorts and subject them to the following analysis: 1) evaluation of mendelian inheritance characteristics, 2) high-resolution reporter gene expression imaging and analysis in embryos and adults, 3) clinical serum chemistry analysis, 4) RNA expression profiling of key tissues and 5) Evaluation of causation and timing of embryonic lethality. Expression analysis of histological specimens involves a semi-automated imaging and analysis workflow. Embryonic phenotypes are analyzed by optical projection tomography (OPT), soft tissue enhanced micro computed tomography (μCT) and 3D image analysis. Data are collected in a custom image and phenotype database. We mine the resulting data in order to understand biological functions of genes of interest as well as to evaluate secreted and transmembrane proteins as potential therapeutic targets for our human antibody-based therapeutic pipeline.

DECODING THE GENOME WITH LIGHT MICROSCOPY

Eugene W Myers

HHMI, Janelia Farm Research Campus, Ashburn, VA, 20147

We are now at a time when we can systematically build protein and promotor constructs across the genome of an organism. Combined with new modalities of light microscopy we can now produce images that allow us to observe molecular mechanisms within the cell, to observe the developmental trajectory of growing organisms, and to map the cellular anatomy of organisms and organs such as the brain, the heart, or the stem of a plant. All this increasingly requires computation to either extract information or to quantitatively measure an effect. This has spawned the sub-field of bioimage informatics.

My group is working on a number of imaging projects along these lines. These include (a) the biophysics of cell division, (b) studies of gene expression in individual cells within the worm *C. elegans*, (c) tracking the whiskers of a mouse while it is learning a response, (d) a detailed reconstruction of a fly's brain including the patterning of its development, and (e) the construction of a high-throughput microscope to image the volume of an entire mouse brain at 1 micron resolution (4.2 trillion voxels) in less than a week. I will touch on all these projects if time permits.

Participant List

Dr. April Agee
Pioneer Hi-Bred International, A DuPont Business
april.agee@pioneer.com

Dr. Erik Andersen
Princeton University
eca@princeton.edu

Dr. Zhirong Bao
Sloan-Kettering Institute
baoz@mskcc.org

Dr. Philip Benfey
Duke University
philip.benfey@duke.edu

Dr. Amy Bernard
Allen Institute for Brain Science
amyb@alleninstitute.org

Dr. John Bixby
University of Miami
jbixby@miami.edu

Dr. Wolfgang Busch
Duke University
wolfgang.busch@duke.edu

Dr. Anne Carpenter
Broad Institute of Harvard and MIT
anne@broadinstitute.org

Dr. Yolanda Chong
University of Toronto
yolanda.chong@utoronto.ca

Dr. Julie Conkright
Stowers Institute for Medical Research
jco@stowers.org

Dr. Christian Conrad
EMBL
cconrad@embl.de

Dr. David Cox
Harvard University
cox@rowland.harvard.edu

Mr. Matthew Crane
Georgia Tech
crane@gatech.edu

Dr. Ian Davis
GrassRoots Biotechnology, Inc.
ian.davis@grassrootsbio.com

Dr. Angela DePace
Harvard Medical School
angela_depace@hms.harvard.edu

Dr. Zhuo Du
Sloan-Kettering Institute
duz@mskcc.org

Prof. Saso Dzeroski
Jozef Stefan Institute
Saso.Dzeroski@ijs.si

Dr. Roland Eils
German Cancer Research Center
r.eils@dkfz.de

Dr. Kevin Eliceiri
University of Wisconsin
eliceiri@wisc.edu

Dr. Suqin Fang
Duke Universysity
sf98@duke.edu

Dr. Bernd Fischer
EMBL
bernd.fischer@embl.de

Dr. Scott Fraser
California Institute of Technology
sefraser@caltech.edu

Mr. Charles Frogner
MIT
frogner@mit.edu

Dr. Nicholas Gale
Regeneron Pharmaceuticals, Inc.
nicholas.gale@regeneron.com

Mr. Taras Galkovskyi
Duke
galkovsky@gmail.com

Dr. Venugopala Gonehal Reddy
Univ. of California, Riverside
vENUG@UCR.EDU

Dr. Sigrun Gustafsdottir
Broad Institute
sigrun@broadinstitute.org

Dr. Omar Gutierrez-Arenas
University of Miami
o.arenas@miami.edu

Dr. Michael Halter
NIST
michael.halter@nist.gov

Mr. Ville Härmä
VTT Technical Research Centre of Finland
ville.harma@vtt.fi

Dr. Joern Hopke
sanofi-aventis
joern.hopke@sanofi-aventis.com

Dr. Wolfgang Huber
EMBL
whuber@embl.de

Dr. Luis Ibanez
Kitware, Inc.
luis.ibanez@kitware.com

Dr. Anjali Iyer-Pascuzzi
Duke University
asi2@duke.edu

Ms. Hueihan Jhuang
MIT
hueihan@mit.edu

Mr. Jee Jung
GrassRoots Biotechnology
jwj2148@hotmail.com

Dr. Philipp Keller
Howard Hughes Medical Institute
kellerp@janelia.hhmi.org

Dr. Melissa Kemp
The Broad Institute
mmkemp@broadinstitute.org

Dr. Imtiazz Khan
Broad Institute of MIT and Harvard
ikhan@broadinstitute.org

Dr. David Knowles
Lawrence Berkeley National Laboratory
dwknowles@lbl.gov

Dr. Judice Koh
University of Toronto
judice.koh@utoronto.ca

Dr. Vance Lemmon
Univ. of Miami
vlemmon@miami.edu

Dr. Xiao Liu
Stanford University
xiaoliu2@stanford.edu

Dr. Vebjorn Ljosa
Broad Institute of MIT and Harvard
ljosa@broad.mit.edu

Mr. Daniel Mace
University of Washington
dlm19@u.washington.edu

Dr. Robert Majovski
Press
majovski@cshl.edu

Mr. Michael Malone
Monsanto Company
michael.malone@monsanto.com

Dr. B. S. Manjunath
University of California
manj@ece.ucsb.edu

Dr. Bradley Martsberger
Duke
btm@duke.edu

Dr. Mojca Mattiazzi Usaj
University of Toronto, Donnelly Centre
m.mattiazziusaj@utoronto.ca

Mr. Michael Mayhew
Duke University
michael.mayhew@duke.edu

Dr. Pauline Menager
CYTOO SA

Dr. Ivar Meyvantsson
Bellbrook Labs
ivarm@bellbrooklabs.com

Dr. Yuriy Mileyko
Duke University
yury@math.duke.edu

Ms. Julia Moore
Sloan-Kettering
moorej@mskcc.org

Dr. John Murray
University of Pennsylvania School of Medicine
jmurr@mail.med.upenn.edu

Dr. Eugene Myers
HHMI
myersg@janelia.hhmi.org

Dr. Matthias Nees
VTT Technical Research Centre of Finland
matthias.nees@vtt.fi

Prof. Uwe Ohler
Duke University
uwe.ohler@duke.edu

Dr. Klaus Palme
University of Freiburg
klaus.palme@biologie.uni-freiburg.de

Dr. Gregoire Pau
EMBL
gregoire.pau@embl.de

Dr. Lucas Pelkmans
Institute of Molecular Systems Biology
pelkmans@imsb.biol.ethz.ch

Dr. Pietro Perona
California Institute of Technology
perona@vision.caltech.edu

Dr. Norbert Perrimon
Harvard Medical School
perrimon@receptor.med.harvard.edu

Dr. Fabio Piano
New York University
fp1@nyu.edu

Dr. Anne Plant
NIST
anne.plant@nist.gov

Dr. Arvind Rao
Carnegie Mellon University
ukarvind@andrew.cmu.edu

Dr. Thomas Sandmann
German Cancer Research Center
t.sandmann@dkfz.de

Dr. anthony Santella
Sloan Kettering Institute
santella@mskcc.org

Dr. Samir Sawant
National Botanical Research Institute
samirsawant@nbri.res.in

Dr. William Schafer
MRC Laboratory of Molecular Biology
wschafer@mrc-lmb.cam.ac.uk

Dr. Alexander Schliep
Rutgers University
schliep@cs.rutgers.edu

Dr. Ethan Schonbrun
Rowland Institute at Harvard University
schonbrun@rowland.harvard.edu

Mr. Adrian Scott
Institute for Systems Biology
awells@systemsbiology.org

Mr. Vasanth Singan
University College Dublin, Ireland
Vasanth.Singan@ucd.ie

Ms. Elizabeth Smith
Merck Research Labs
elizabeth.smith@spcorp.com

Dr. Iman Soliman
AinShams Fac of Medicine
iman1205@gawab.com

Mr. Christoph Sommer
University of Heidelberg
christoph.sommer@iwr.uni-heidelberg.de

Ms. Pranidhi Sood
New York University
psood@nyu.edu

Prof. Peter Sorger
Harvard Medical School
psorger@mit.edu

Dr. Edgar Spalding
University of Wisconsin
spalding@wisc.edu

Dr. Ernst Stelzer
EMBL (European Molecular Biology Lab)
stelzer@embl-heidelberg.de

Dr. Nico Stuurman
Biozentrum der Universitat Basel
STUURMAN@UBACLU.UNIBAS.CH

Ms. Erin Styles
Donnelly CCBR, University of Toronto
erin.styles@utoronto.ca

Dr. Hillary Sussman
Genome Research
hsussman@cshl.edu

Dr. Jason Swedlow
University of Dundee
jason@lifesci.dundee.ac.uk

Dr. Edmund Talley
National Institutes of Health
talleye@mail.nih.gov

Ms. Andreia Teixeira-Castro
University of Minho
accastro@ecsaude.uminho.pt

Dr. Pavel Tomancak
MPI-CBG
tomancak@mpi-cbg.de

Dr. Srinivas Turaga
MIT
sturaga@mit.edu

Dr. Pierre Turpin
Molecular Devices
pierre.turpin@moldev.com

Dr. Maria Cristina Ubach
Monsanto Company
cristina.ubach@monsanto.com

Dr. Joerg Vandenhirtz
LemnaTec TmbH
joerg.vandenhirtz@lemnatec.de

Prof. Carolina Wahlby
Broad Institute
carolina@broadinstitute.org

Dr. Thomas Walter
EMBL
twalter@embl.de

Dr. Meng Wang
Monsanto Company
meng.wang@monsanto.com

Prof. Xiujie Wang
Institute of Genetics & Developmental Biology, CAS
xjwang@genetics.ac.cn

Mr. Quanli Wang
Duke University
quanli@stat.duke.edu

Dr. Robert Waterston
University of Washignton
watersto@u.washington.edu

Ms. Amelia White
New York University
white.amelia@gmail.com

Dr. Winfried Wiegraebe
Stowers Institute for Medical Research
wiw@stowers.org

Mr. Michael Wong
Mouse Imaging Centre
mwong@phenogenomics.ca

Prof. Lani Wu
UT Southwestern Medical Center
lani.wu@utsouthwestern.edu

Prof. Mehmet Yanik
MIT
yanik@mit.edu

Dr. John Yordy
M. D. Anderson Cancer Center
jsyordy@mdanderson.org

Dr. Joanne Young
CYTOO SA
jyoung@cytoo.com

Mr. Lee Zamparo
University of Toronto
zamparo@cs.toronto.edu

Dr. Lifang Zhang
Cold Spring Harbor Laboratory
zhangl@cshl.edu

Dr. Jingxin Zhang
sanofi-aventis
jingxin.zhang@sanofi-aventis.com

Dr. Xian Zhang
German Cancer Research Center
x.zhang@dkfz.de

VISITOR INFORMATION

EMERGENCY	CSHL	BANBURY
Fire	**(9) 742-3300**	**(9) 692-4747**
Ambulance	**(9) 742-3300**	**(9) 692-4747**
Poison	**(9) 542-2323**	**(9) 542-2323**
Police	**(9) 911**	**(9) 549-8800**
Safety-Security	**Extension 8870**	

Emergency Room **Huntington Hospital** 270 Park Avenue, Huntington	**631-351-2300** **(1037)**
Dentists Dr. William Berg Dr. Robert Zeman	 **631-271-2310** **631-271-8090**
Doctor MediCenter 234 W. Jericho Tpke., Huntington Station	**631-423-5400** (**1034**)
Drugs - 24 hours, 7 days Rite-Aid 391 W. Main Street, Huntington	**631-549-9400** (**1039**)

Free Speed Dial
Dial the four numbers (********) from any **tan house phone** to place a free call.

GENERAL INFORMATION

Books, Gifts, Snacks, Clothing, Newspapers
BOOKSTORE 367-8837 (hours posted on door)
Located in Grace Auditorium, lower level.

Photocopiers, Journals, Periodicals, Books, Newspapers
Photocopying – Main Library
Hours: 8:00 a.m. – 9:00 p.m. Mon-Fri
10:00 a.m. – 6:00 p.m. Saturday
Helpful tips - Obtain PIN from Meetings & Courses Office to enter Library after hours. See Library staff for photocopier code.

Computers, E-mail, Internet access
Grace Auditorium
Upper level: E-mail only
Lower level: Word processing and printing.
STMP server address: mail.optonline.net
To access your E-mail, you must know the name of your home server.

Dining, Bar
Blackford Hall
Breakfast 7:30–9:00, Lunch 11:30–1:30, Dinner 5:30–7:00
Bar 5:00 p.m. until late
Helpful tip - If there is a line at the upper dining area, try the lower dining room

Messages, Mail, Faxes

Message Board, Grace, lower level

Swimming, Tennis, Jogging, Hiking

June–Sept. Lifeguard on duty at the beach. 12:00 noon–6:00 p.m. Two tennis courts open daily.

Russell Fitness Center

Dolan Hall, east wing, lower level
PIN#: **Press 64575 (then enter #)**

Concierge

On duty daily at Meetings & Courses Office.
After hours – From tan house phones, dial x8870 for assistance

Pay Phones, House Phones

Grace, lower level; Cabin Complex; Blackford Hall; Dolan Hall, foyer

CSHL's Green Campus

Cold Spring Harbor Laboratory is pledged to operate in an environmentally responsible fashion wherever possible. In the past, we have removed underground oil tanks, remediated asbestos in historic buildings, and taken substantial measures to ensure the pristine quality of the waters of the harbor. Water used for irrigation comes from natural springs and wells on the property itself. Lawns, trees, and planting beds are managed organically whenever possible. And trees are planted to replace those felled for construction projects.

Two areas in which the Laboratory has focused recent efforts have been those of waste management and energy conservation. The Laboratory currently recycles most waste. Scrap metal, electronics, construction debris, batteries, fluorescent light bulbs, toner cartridges, and waste oil are all recycled. For general waste, the Laboratory uses a "single stream waste management" system, removing recyclable materials and sending the remaining combustible trash to a cogeneration plant where it is burned to provide electricity, an approach considered among the most energy efficient, while providing a high yield of recyclable materials.

Equal attention has been paid to energy conservation. Most lighting fixtures have been replaced with high efficiency fluorescent fixtures, and thousands of incandescent bulbs throughout campus have been replaced with compact fluorescents. The Laboratory has also embarked on a project that will replace all building management systems on campus, reducing heating and cooling costs by as much as twenty-five per cent.

Cold Spring Harbor Laboratory continues to explore new ways in which we can reduce our environmental footprint, including encouraging our visitors and employees to use reusable containers, conserve energy, and suggest areas in which the Laboratory's efforts can be improved. This book, for example, is printed on recycled paper.

1-800 Access Numbers

AT&T	**9-1-800-321-0288**
MCI	**9-1-800-674-7000**

Local Interest

Fish Hatchery	631-692-6768
Sagamore Hill	516-922-4447
Whaling Museum	631-367-3418
Heckscher Museum	631-351-3250
CSHL DNA Learning Center	x 5170

New York City

Helpful tip -

Take Syosset Taxi to Syosset Train Station ($8.00 per person, 15 minute ride), then catch Long Island Railroad to Penn Station (33rd Street & 7th Avenue). Train ride about one hour.

TRANSPORTATION

Limo, Taxi

Syosset Limousine	516-364-9681 (**1031**)
Super Shuttle	800-957-4533 (**1033**)
To head west of CSHL - Syosset train station	
Syosset Taxi	516-921-2141 (**1030**)
To head east of CSHL - Huntington Village	
Orange & White Taxi	631-271-3600 (**1032**)
Executive Limo	631-696-8000 (**1047**)

Trains

Long Island Rail Road	822-LIRR
Schedules available from the Meetings & Courses Office.	
Amtrak	800-872-7245
MetroNorth	800-638-7646
New Jersey Transit	201-762-5100

Ferries

Bridgeport / Port Jefferson	631-473-0286 **(1036)**
Orient Point/ New London	631-323-2525 **(1038)**

Car Rentals

Avis	631-271-9300
Enterprise	631-424-8300
Hertz	631-427-6106

Airlines

American	800-433-7300
America West	800-237-9292
British Airways	800-247-9297
Continental	800-525-0280
Delta	800-221-1212
Japan Airlines	800-525-3663
Jet Blue	800-538-2583
KLM	800-374-7747
Lufthansa	800-645-3880
Northwest	800-225-2525
United	800-241-6522
US Airways	800-428-4322